Facultad de Ciencias

Departamento de Biologia Molecular

Structural and functional description of the mouse *Tyr* locus

Davide Seruggia

Madrid, 2014

Facultad de Ciencias

Departamento de Biologia Molecular

Structural and functional description of the mouse *Tyr* locus

A thesis submitted in fulfilment of the request to obtain the PhD degree in Biochemistry, Molecular Biology, Biomedicine and Biotechnology at the Universidad Autonoma de Madrid by

Davide Seruggia

Degree in Biotechnology by University of Milano-Bicocca

PhD Director: **Dr Lluis Montoliu Jose**

Laboratory ''Animal models by genetic manipulation''

Department of Molecular and Cellular Biology

National Centre for Biotechnology (CNB)

State Agency Spanish National Research Council (CSIC)

Universidad Autonoma de Madrid

Madrid, 2014

This PhD thesis has been carried out by Davide Seruggia at the ‘’Animal Models by Genetic Manipulation ‘’ laboratory from the Dept. of Molecular and Cellular Biology at the National Centre for Biotechnology (CNB-CSIC), in Madrid, under the supervision of Dr. Lluis Montoliu Jose, CSIC Research Scientist, CIBER-ER (Network Centre of Research in Biomedicine on Rare Diseases), ISCIII (Instituto de Salud Carlos III) Researcher.

The support received from the following Grants and Fellowships has permitted to develop this PhD work:

- La Caixa International PhD Fellowship Program, 2010 Call. Recipient: Davide Seruggia (2010-2014).

- EMBO Short Term Fellowship, ASTF 140-2013. Recipient: Davide Seruggia.

- Functional and structural validation of genomic boundaries. Spanish Ministry of Economy and Competitiveness (MINECO) National Plan R+D+I, Biotechnology Program, 2013-2015, reference BIO2012-3998D. PI: Dr. Llius Montoliu Jose.

- Identification and functional validation of genomic vertebrate insulators, *in vitro* and *in vivo*, Spanish Ministry of Science and Innovation, National Plan R+D+I, Biotechnology Program, 2009-2012, reference BIO2009-12697. PI: Dr. Llius Montoliu Jose.

- EMMAservice: Servicing the European Biomedical Research Community: archiving and dissemination of Mouse Models of Human Disease, European Commission, FP7-intrastructures-2008-1, 2009-2012, reference ID 227490. PI: Dr. Lluis Montoliu Jose (ten partners, coordinator: Prof Glauco Tocchini-Valentini, CNR-IBC, Monterotondo, Rome, Italy).

- Genetic diagnosis and potential therapies for albinism, CIBER-ER, ISCIII, Spanish Ministry of Economy and Competitiveness (MINECO), 2012-2014, reference 13-756/152.09. PI: Dr. Lluis Montoliu Jose, coordinator (one additional group).

- Animal models for investigating disease of the visual system, R+D Programs in-

Biomedicine, Autonomous Government of Madrid, 2012-2015, reference S2011/BMD-2439. PI Dr. Lluis Montoliu Jose, coordinator (four other groups).

- INFRAFRONTIER-I3: Development of mouse mutant resources for functional analysis of human diseases – Enhancing the translation of research into innovation, European Commission, FP7 Capacities, 2013-2016, INFRA-2012-1.1.4, reference 212325. PI: Dr. Lluis Montoliu Jose (23 partners, coordinator: Prof. Dr. Martin Hrabe de Angelis, IEG-HMGU, Munich, Germany),

- International Society for Transgenic Technologies (ISTT), Registration Award to attend the 12th Transgenic Technology (TT2014) Meeting held in Edinburgh, UK. Recipient: Davide Seruggia.

- Spanish Society of Biochemistry and Molecular Biology (SEBBM), Registration Award to attend the SEBBM Meeting held in Barcelona, Spain, 2011. Recipient: Davide Seruggia.

- International Mammalian Genome Society (IMGS), Registration Award to attend the IMGS Meeting, held in Salamanca, Spain, 2013. Recipient: Davide Seruggia.

- University of Cantabria (UIMP/UC), Registration Award to attend the Cantabria Campus Nobel Meeting, held in Santander, Spain, 2012. Recipient: Davide Seruggia.

ACKNOWLEDGMENTS

My first acknowledgments go to Dr. Lluis Montoliu, for supporting my work in the laboratory and for directing my PhD work. I would like to especially thank you for the freedom that I had during these years, to design and perform a number of experiments, not always exactly *on theme*. From this *parallel* work, I learn to measure my skills and limits and tasted the highs and lows of independent research.

A big thank you goes to Dr. Pawel Pelczar that hosted me in the laboratory during my stay in Zurich. I really learn a lot in such a short period of time.

I would like to acknowledge all present and past member of the lab 111 for the good (and bad!) time we have shared. I have to acknowledge the *Andalusian girls* Almudena and Cristina who suffered my alternating mood as well as Pau and Irene who experienced the pain of having me as a boss.

SUMMARY

The complete set of regulatory elements that are necessary to achieve correct gene expression – including enhancers, silencers and chromatin insulators – can be spread in the range of kilo bases to mega bases, far away from the corresponding coding sequence. Description of such complex picture requires efforts in the identification of key elements, from a functional and structural point of view, using reporter assays in cultured cells and trying to recapitulate gene expression – artificially – in transgenic mice. In this work, we describe our experiments aimed to unravel the structure and the functioning of the mouse *Tyr* locus. We focused our attention on chromatin boundaries, starting from previous results of the laboratory. By Chromosome Conformation Capture (3C), the *Tyr* 5' boundary (or LCR, locus control region) has been characterized. Next, a second boundary, found on the 3' end of the locus, was identified, tested *in vitro* and *in vivo* and finally – by 3C – found to be in physical contact the *Tyr* promoter in all the cell types analysed. Two different mechanisms – CTCF occupancy and SINEB2 polymerase II/III transcription – confer insulator property to this new boundary, which is also engaged in conferring a 3D structure to the *Tyr* locus. Finally, to prove the *in vivo* relevance of the two boundary elements described in this study, several transgenic mouse lines were generated carrying the inactivation of each of these sequences, using the new CRISPR/Cas9 system. Homozygous individuals carrying the inactivation of the *Tyr* 5' boundary present a reduction in coat colour. Coat pigmentation is not fully absent – as in albino mice – but is strongly reduced, indicating that the *Tyr* 5' element is relevant for coordinating *Tyr* expression. Finally, transgenic mouse lines carrying the inactivation of the *Tyr* 3' boundary in albino background did not show alteration in coat colour pigmentation, suggesting that this second conserved element is not required for the control of *Tyr* expression, and might be necessary to control the flanking *Nox4* gene.

PRESENTACIÓN

El juego completo de los elementos reguladores que son necesarios para lograr la expresión correcta de los genes, lo cual incluye potenciadores, silenciadores y fronteras genómicas, puede estar distribuido a lo largo de muchos centenares de kilobases, muy alejados de la secuencia que codifica para la proteína. La descripción de este paisaje complejo requiere esfuerzos en la identificación de elementos clave, desde un punto de vista funcional y estructural, mediante el uso de ensayos de elementos indicadores en células en cultivo, e intentando reproducir la expresión génica de interés, artificialmente, en ratones transgénicos. En este trabajo describimos experimentos encaminados a descifrar la estructura y la función del locus *Tyr* del genoma de ratón. Hemos focalizado nuestra atención en las fronteras genómicas, a partir de resultados previos del laboratorio. Mediante experimentos de Captura de Conformación Cromosomal (3C), la frontera *Tyr* 5' (también conocida como LCR, de Región Controladora del Locus), ha sido caracterizada. A continuación, una segunda frontera genómica, situada en la región 3' del locus, ha podido ser identificada, probada in vitro e in vivo, y, finalmente, mediante 3C, se ha constatado su contacto físico con el promotor del gen *Tyr* en todos los tipos celulares analizados. Sabemos que dos mecanismos diferentes (la presencia de CTCF y la transcripción a partir de los promotores RNA polimerasa II/III de los elementos SINEB2) confieren la propiedad aisladora a esta nueva frontera genómica, la cual está igualmente involucrada en la estructura tridimensional del locus *Tyr*. Seguidamente, para probar la importancia, in vivo, de estas dos fronteras genómicas descritas en este trabajo, hemos generado diversas líneas de ratones transgénicos con mutaciones en cada una de estas secuencias, utilizando el nuevo sistema CRISPR/Cas9. Los individuos homocigotos que son portadores de la mutación en el elemento *Tyr* 5' muestran una reducción en la pigmentación del pelaje. Esta

pigmentación no desaparece por completo, como ocurre en el caso de los ratones albinos, pero sí que se reduce de forma muy significativa. Estos resultados indican que el elemento *Tyr* 5' es relevante para la expresión correcta del gen *Tyr*. Finalmente, las líneas de ratones transgénicos portadoras de mutaciones en el elemento *Tyr* 3' no mostraron alteración aparente en la pigmentación del pelaje, en fondo albino, lo cual sugiere que este segundo elemento regulador, conservado evolutivamente, no es indispensable para el control de la expresión del gen *Tyr*, pero sin embargo podría ser necesario para el control del gen vecino *Nox4*.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3C	chromosome conformation capture
BAC	bacterial artificial chromsome
bp	base pair
CAR	cardiac actin promoter
Cas9	crispr-associated protein 9
ChIp-seq	Chromatin Immunoprecipitation followed by sequencing
CMV	CytoMegalo Virus
CNS	central nervous system
CRISPR	clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
DNAseI	Deoxyribonuclease I
EDTA	Ethylene Diamine Tetraacetic Acid
ENCODE	ENCyclopedia of DNA elements
ES cells	Embryonic Stem cells
F1	First filial generation of transgenic mice
FBS	Fetal Bovine Serum
GFP	Green Fluorescent protein
HEK 293	Human Embryonic Kidney 293 cells
HEPES	Hydroxy Ehyl Piperazine Ethane Sulfonic acid
hpf	hours post fertilization
HS	hypersensitive site (to DNAse I cleavage)
Kb	kilobase
LB	Luria-Bertani broth
LCR	Locus control region
Mb	mega base pair
mq H_20	milli Q water
mRNA	messenger RNA
o/n	overnight
ONPG	Ortho-Nitrophenyl-b-galactoside
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	paraFormaldehyde

PTU	1-phenyl methane sulfonyl fluoride
qPCR	quantitative polymerase chain raction
RNA	ribonucleic acid
RT	room temperature
SD	standard deviation
SEM	standar error of the mean
SINE	short interspersed nuclear element
SV40	simian vacuolating virus 40
TAE	tris acetate-EDTA
TALEN	transcription activator-like effector nuclease
Taq	thermus aquaticus
TE	tris-EDTA
Tm	melting temperature
ZFN	zinc finger nuclease

1. Introduction

1.1 Mouse *Tyr* and its regulatory elements

The tyrosinase gene encodes the key enzyme in the melanin production pathway, a copper-containing di-oxidase that triggers the formation of melanin. It catalyses the rate-limiting conversion of L-tyrosine to DOPAquinone in two consecutive oxidazing steps. In mice, *Tyr* is located in chromosome 7 (Kwon et al. 1989) and it is composed of 5 coding exons and 4 introns spanning a 66 kb large region. One main isoform is known to be translated into a protein (Muller et al. 1988), 533 residues long; smaller transcripts are generated by alternative splicing (Ruppert et al. 1988). *Tyr* is expressed in neural crest-derived melanocytes of skin, choroid, hair follicles, iris, inner ear and, among other places in the body, the heart (Carneiro and Kos 2008) and in the retinal pigmented epithelium of the retina (RPE), which derives from the outer wall of the optic cup (Beermann et al. 1992, Gimenez et al. 2003). Mutations in the tyrosinase coding sequence are responsible for the oculocutaneous albinism type I (OCA1, OMIM 203100, Montoliu et al. 2014). In mice, the causative albino mutation appears to be a change of guanine 390 to cytosine, resulting in a change of cysteine 103 to serine (Kwon et al., 1988; Halaban et al. 1988), named Tyr^{c}. This amino acid substitution alters the copper-binding site and results in complete loss of enzymatic activity. Jackson and Bennett (1990) reported the molecular basis of albinism. Interestingly, more than 20% of OCA1 cases can't be associated with mutations in the *Tyr* coding sequence, thus suggesting that mutations inactivating the locus can also occur in non-coding regulatory elements (King et al. 2003; Montoliu et al. 2014).

1.1.1 Proximal *Tyr* regulatory elements

The mouse *Tyr* promoter – along with that of other species – has been extensively characterized (Ganss et al. 1994; Ferguson and Kidson, 1997). Although a canonical TATA-binding site is absent in both mouse and human promoters, TATA-like sequences could be identified at approximately 32 bp upstream the TSS (Ruppert et al. 1988; Yamamoto et al. 1989). In correspondence with the TSS, a DNAse I-resistant element was found in the human TYR promoter and named *initiator* (*Inr*, **Figure i1**).

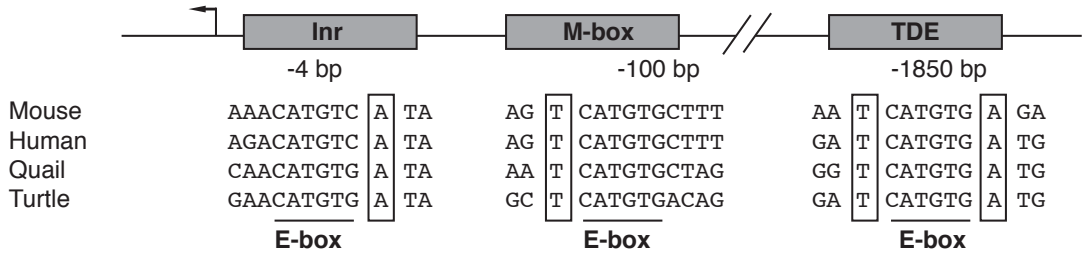


Figure i1 Conserved motives in the *tyrosinase* promoter. Two immediate proximal elements and one distal element are found in the mouse Tyr upstream sequence that are conserved in many orthologs. These three motives can be occupied by bHLH-ZIf transciption factors, such as MITF. Adapted from Bentley et al. 1994.

The *Inr*, conserved in mouse, chicken, quail and turtle, contains an E-box consensus motif (CANNTG) that is essential for human promoter function (Bentley et al. 1994). The Inr E-box is recognized by two helix-loop-helix leucine zipper transcription factors (bHLH-ZIP), namely Upstream stimulatory factor (USF) and Microphthalmia-associated transcription factor (MITF), a transcription factor essential for pigment cells development (Goding, 2000). In fact, homozygous mutations in the *Mitf* gene can result in non-pigmented mice that develop small, unpigmented eyes (Hodgekinson et al. 1993; Arnheiter, 2010). In search of pigment-cell specific factors, the *Tyr* promoter sequence was compared with that of paralog genes, including *Tyrp1* and *Dct*, in different species. A conserved, 11 bp motif was found between 40 and 160 bp upstream the TSS, termed M-box (Lowings et al. 1992), in several orthologs. Similarly to the E-box, the M-box contains a CANNTG motif and can be bound by MITF. An additional, tissue-specific enhancer element was found 1.8 kb upstream the TSS in the human TYR promoter, conserved in mouse (Shibata et al. 1992). This element, named Tyrosinase distal element (TDE) also contains an E-box (Yasumoto et al. 1994) and can be trans-activated by MITF. Gene transfer experiments performed by independent groups in the early 90's showed that indeed such promoter elements are able to rescue the pigmented phenotype when used to drive the expression of a tyrosinase minigene in albino mice (Beerman et al. 1990; Kluppel et al. 1991). Partial rescue of pigmentation could be obtained using different minigene constructs including a 5.5 kb fragment 5' upstream the TSS as promoter (Beerman et al. 1990), a 2.6 kb fragment (Tanaka et al. 1990), a 2.2 kb fragment (Yokoyama et al. 1990), as well as minimal, 270 bp fragment able to drive weak and variable transgene expression in the retinal pigmented epithelium and in skin melanocytes (Kluppel et al. 1991, Beermann et al. 1992; Ganss et al. 1994). Transgenic mice showed a variable degree of pigmentation, not reaching

the wild-type level, due to chromosomal position effects (Wilson et al. 1990; Giraldo and Montoliu 2002), and because additional, upstream regulatory elements are necessary in order to reproduce the expression level of the endogenous gene.

1.1.2 Distal *Tyr* regulatory elements

The use of small constructs carrying minimal regulatory sequences helped to formulate fundamental biological statements, such as that the albino phenotype can be rescued by a functional tyrosinase transgene. However, to define all the regulatory elements of a locus, minigene constructs are definitely too short. Hence, larger constructs carrying all distal regulatory elements are necessary. Thanks to technical advances in the production of transgenic animals using yeast artificial chromosomes (YACs) (Schedl et al. 1992), larger DNA constructs could be used for to study the role of distal regulatory elements. Schedl, Montoliu and colleagues introduced in albino mouse fertilized eggs a 250 kb YAC containing 80 kb of the mouse *Tyrosinase* coding region and 155 kb of upstream sequences (YRT2 YAC, Schedl and Montoliu 1993). The use of such a large construct allowed full rescue of the albino phenotype in all transgenic mice obtained, that are phenotypically indistinguishable to pigmented, wild-type individuals (**Figure i2**).

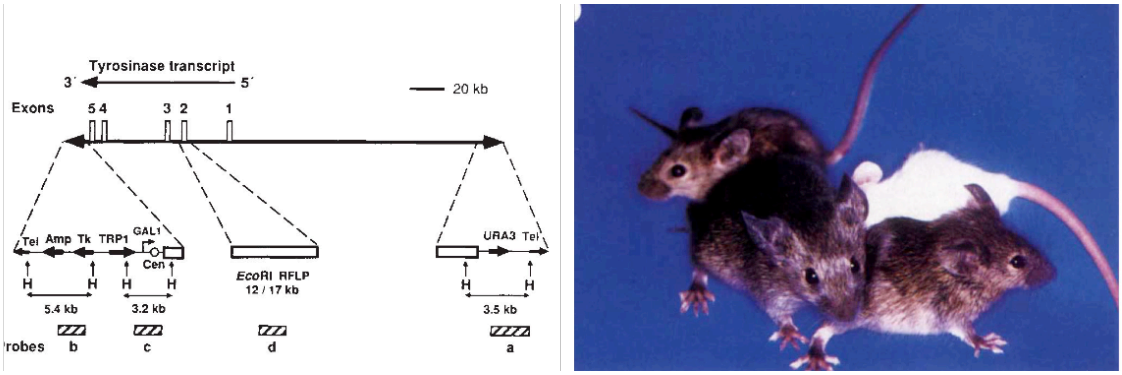


Figure i2 Full rescue of pigmentation using the YRT2 YAC. **Right** Scheme of the YRT2 YAC and position of the probes used to detect the integrity of the YAC in transgenic mice by Southern Blot. **Left** Three F1 individuals from three independent transgenic lines showing indistinguisble, pigmented phenotype, compared with a non-transgenic littermate. Adapted from Schedl et al. 1993

Using large, genomic-type constructs, copy number-dependent and position-independent expression of the transgene could be achieved (Giraldo and Montoliu, 2001). The observed differences with respect to the variegated *Tyr*

expression obtained using minigene vectors depends uniquely on distal *Tyr* regulatory elements contained in the 155 kb non-coding DNA included in the YRT2 YAC. This observation opened a quest aimed to the *cis*-acting element responsible for the successful rescue of pigmentation in the YRT2 transgenic mouse line (reviewed in Giraldo and Montoliu, 2002). The nature of the distal, yet unknown element is discussed in the original paper (Schedl et al. 1993). Hypothesis were formulated whether a dominant Locus control region (LCR) was found, similar to what had been reported using a human β-globin large transgene for generating a transgenic mouse (Grosveld et al. 1987) or in gene transfer experiments using elements from the chicken lysozyme gene (Stief at al. 1989).

Molecular analysis of a spontaneous mutant mouse, the *chinchilla-mottled* mouse (*Tyr*$^{c\text{-}m}$), was instrumental in the identification of the *Tyr* distal upstream element (Porter et al. 1991). Melanocytes from *Tyr*$^{c\text{-}m}$ mice showed reduced *Tyr* expression and distinct DNAse I hypersensitivity with respect of melanocytes from wild-type mice. In particular, a hypersensitive site in melanocytes obtained from wild-type mice found 15 kb upstream the *Tyr* TSS was missing in *Tyr*$^{c\text{-}m}$ melanocytes. This region underwent chromosomal rearrangement in *Tyr*$^{c\text{-}m}$ mouse likely resulting in the observed reduction and striped-like pigmentation phenotype. Comparing DNAse I hypersensitivity of the *Tyr* upstream sequence in distinct mouse cell lines, a cell-type specific DNAse I HS site was found in B16 mouse melanoma cells (**Figure i3**) (Ganss et al. 1994).

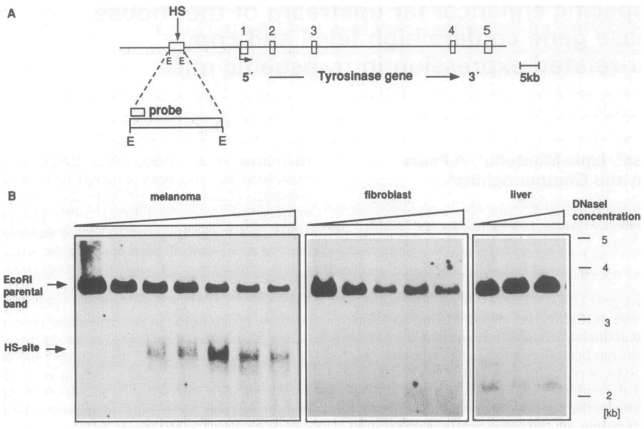


Figure i3 DNAse I hypersensitivity of *Tyr* upstream sequences. **A** Position of the Tyr locus along with the southern blot probe used for the DNAse hypersentivity assay. **B** A cell-type specific HS site was identified 15 kb upstream the Tyr conding sequence. The same HS signal could not be detected in fibroblasts nor in hepatocytes. Adapted from Ganss et al 1994.

The inclusion of such *Tyr* -15 kb HS element into *tyrosinase* minigene transgenes fully rescued the albino phenotype in transgenic mice (**Figure i4**), but also conferred position-independent and copy number-dependent expression of the

transgene (Ganss et al. 1994; Porter and Meyer, 1994). For this reason, this element is referred as the *Tyr* Locus Control Region (thereafter, LCR). In addition, the *Tyr* LCR acts as enhancer in transient transfection in human and mouse melanoma cells (Ganss et al. 1994; Giraldo et al. 2003). The key element contained in the 155 kb non-coding DNA of YAC YRT2 was therefore identified.

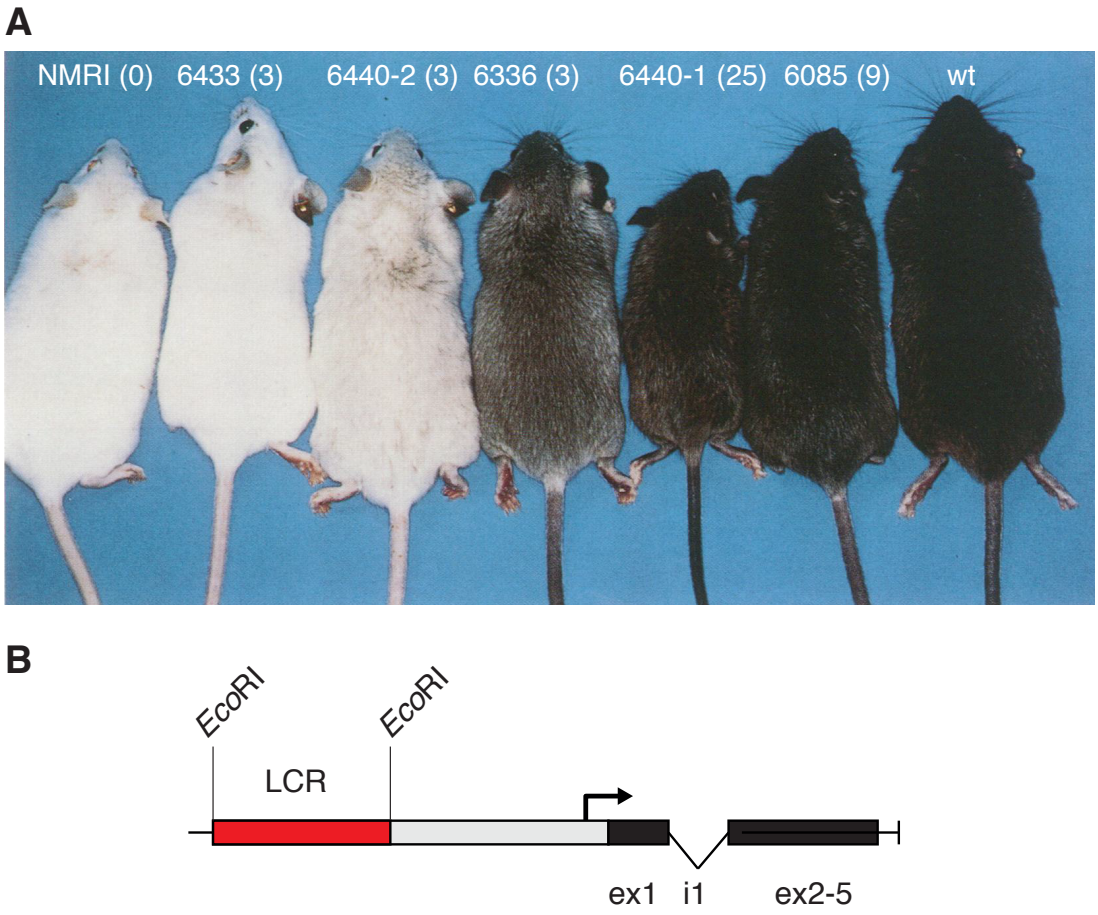


Figure i4 *Tyr* HS element in transgenic mice. **A** Coat colour of transgenic F1 offspring carrying a mini *Tyr* gene coupled with the -15 kb HS element (LCR). Mice are ordered with respect to the degree of pigmentation. An albino, NMRI mouse is shown as control at the left. Similarly a wild-type agouti mouse is shown at right. Estimation of transgene copy number obtained by Southern Blot, indicated between brackets for each transgenic line. **B** Scheme of the phsTyr4 transgene used to generate the transgenic animals in A. A 3.7 kb *Eco*RI fragment of the Tyr 5' upstream region is coupled with a 5.5 kb fragment upstream the *Tyr* coding region and a *Tyr* minigene including the first intron. Adapted from Ganss et al. 1994.

The nature of the *Tyr* LCR was extensively characterized *in vitro* and *in vivo.* The cell type-specific enhancer activity was confirmed *in vitro* and found to be associated with the binding of the transcription factors CREB (cAMP responsive element binding) and AP-1 (activator protein 1), whose binding motives were identified in two fragments named A and B boxes, respectively (Ganss et al. 1994). The putative boundary activity of the *Tyr* LCR was studied in transgenic

flies and mice (Giraldo et al. 2003) and *in vitro* (Angel Garcia Diaz, PhD Thesis, 2007 and Lucia Regales, PhD thesis 2005).

In mice, the inclusion of the 3.7 kb LCR fragment rescued the expression of a luciferase reporter cassette driven by the HSV-TK promoter, which would be otherwise subjected to silencing and variegation because of chromosomal position effect, indicating that such sequence acts as chromatin boundary. Similarly, using the *white* minigene assay in drosophila (Kellum and Schedl, 1991), the barrier activity of the *Tyr* LCR was demonstrated (Giraldo et al. 2003) in transgenic flies. Thee element were found to be linked with the *Tyr* 5' insulator activity: the A and B boxes, bound by CREB and AP-1 and G-rich box, a simple repeat of 15 G nucleotides located at the boundary with a heavily methylated *in vivo*, LINE1 repeat element. Although the sum of the elements conveys a strong insulator property (Bessa et al. 2009), the contribution of each element – taken individually – has not been explored in sufficient detail, hence suggesting a complex mechanism of action.

A second regulatory element was described 48 kb 5' upstream the *Tyr* gene. Using a BAC carrying 186 kb of the mouse *Tyr* locus (including the *Tyr* coding region and 86 kb of 5' upstream non-coding sequences) the rescue of the albino phenotype could be obtained. A series of *lacZ* large constructs based of this BAC were generated carrying deletions in region 5' upstream to the *Tyr* LCR. Interestingly, a non-coding element that seems to be relevant for the expression of *Tyr* in the RPE was identified using transient BAC transgenesis (Murisier et al., 2007). However, the analysis was only carried in transient transgenic mice, without establishing lines, and the number of transient transgenic embryos analysed was very limited (n = 7), not constituting statistically significant data. Additionally, the use of *lacZ* as a reporter gene could be source of confusion (Rosenberg et al. 1992; Montoliu et al. 2000; Montoliu et al. 2004; Bolon et al., 2008).

The presence of an additional boundary element at the 3' end of the locus was hypothetized in our laboratory (Angel Garcia Diaz thesis, 2007; Molto et al. 2009) although this hypothesis hasn't yet been supported by experimental data.

1.2 Appoaches for the inactivation of *Tyr cis*-acting elements *in vivo*

1.2.1 Approximations using large transgenes

Inactivation of a fundamental *cis*-acting regulatory element *in vivo* is expected to produce evident alteration in the expression of the target associated gene or genes. If no differences in gene expression are observed in the absence of the element in study, this is probably dispensable or secondary. Several laboratories achieved the inactivation of regulatory elements using large transgene constructs (BACs and YACs), coupled to a reporter gene, or carrying endogenous gene whose activity can be easily measured – in the absence of the endogenous gene product. Using recombineering techniques in yeast (Reeves et al. 1992; Schedl et al. 1992; Giraldo et al. 1999) and in bacteria (Copeland et al. 2001), the relevant *cis*-acting sequence is deleted or mutated within a large stretch of genomic sequence of the locus. Transgenic animals are generated using such constructs and their phenotype is compared with that of control animals carrying the intact large transgene (Montoliu et al 1996; Peterson et al. 1996, 1998; Gimenez et al. 2001; Gimenez et al. 2005; Horvath et al. 2013; van den Boogaard et al. 2014).

To investigate whether the *Tyr* LCR is the main player in the regulation of *Tyr* expression – or if additional elements are required to coordinate gene expression in distinct tissues and developmental stages, a series of deletions and other genetic modifications based on the original YRT2 YAC (Schedl et al. 1993) were produced (**Figure i5**) and the corresponding YACs were used to generate transgenic mouse lines (**Figure i6**).

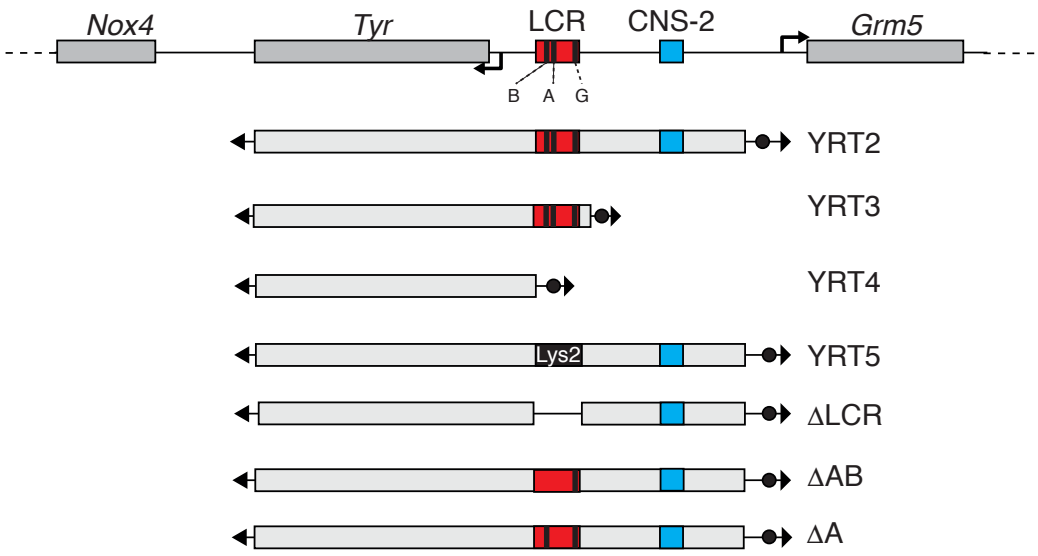


Figure i5 YRT2 YAC and its derivatives constructs. The *Tyr* LCR is shown as a red box. Black bars indicates the A,B and G boxes (Giraldo et al. 2003). The CNS-2 element described in Murisier et al. 2007 is shown as a blue box. YRT2 (250kb) carries the entire *Tyr* expression domain. YRT3, 100 kb YAC with a deletion of 5' LCR upstream sequences. YRT4, 100 kb YAC with a deletion of LCR and its upstream sequences. YRT5, 250 kb YAC carrying a targeted deletion of the LCR, substituting A and B core sequences by a yeast selectable marker (*lys2*). Also, transgenic animals were obtained using a YAC carrying individual deletion of the A box, the B box, or both sequences. Telomers are indicated as arrowheads, centromers as circles.

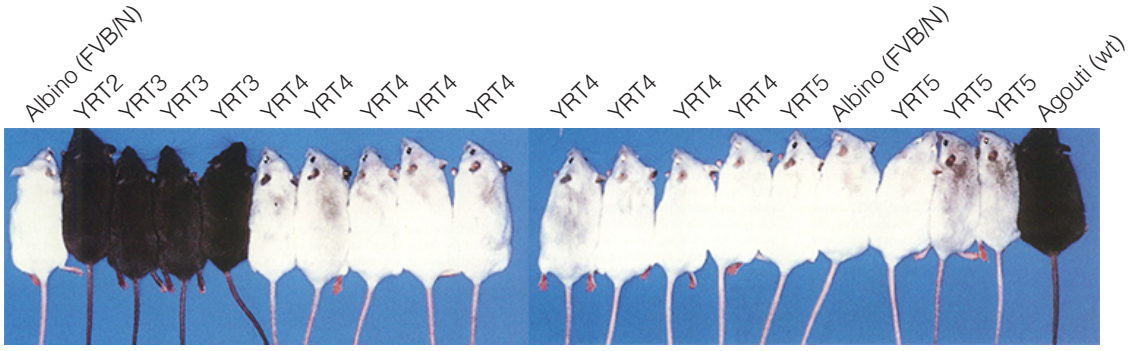


Figure i6 Transgenic mice generated with different *Tyr* YACs. Representative adult individuals from each line obtained with some of the YAC constructs shown in Figure i5. A FVB/N is shown as albino control. A YRT2 and agouti mice serve as wild-type controls. Adapted from Montoliu et al. 1996

Results indicated that the rescue of the albino phenotype correlates strictly with the presence of the *Tyr* LCR (Montoliu et al. 1996): the *Tyr* LCR guarantees position independent and copy number-dependent *Tyr* expression, in both neural crest-derived melanocytes and cells of the retinal pigmented epithelium. In fact, removing the 140 kb of genomic DNA upstream the LCR (in YAC YRT3), leaving the latter intact, did not rendered the *tyrosinase* transgene affected by position effects, indicating that no additional relevant regulatory elements are contained in

the deleted region. In contrast, experiments using engineered BAC-based reporter constructs, indicated that additional regulatory elements might lay upstream the LCR (Murisier et al. 2007). These experiments, however, were performed using transient transgenic lines and are based on the analysis of a very small number of embryos, thus not being stastistically significant, hence suggesting the limited relevance, if any, of the additional regulatory element.

Additional evidences support the central role of the *Tyr* LCR: replacing the LCR sequence with the *Lys2* yeast marker gene in the YRT5 YAC, does affect the correct transgene expression, making it vulnerable to position effects. YACs carrying additional modifications were also produced in our laboratory. By homologous recombination in yeast, the YRT2 YAC was modified in order to completely remove the LCR sequence in the absence of selection markers (differently form the YRT5 YAC which carries the yeast *Lys2* selection marker gene instead) or other leftover sequences, and transgenic mice were obtained (Moreira et al., 2004). Similarly, small 20-60 bp deletions were introduced into the YRT2 YAC sequence, targeting two key elements contained in the *Tyr* LCR, the A and B boxes (Giraldo et al., 2003) (**Figure i5**). Transgenic mice were obtained from YRT2ΔA and YRT2ΔAB YACs that showed a reduction of coat pigmentation. However, the observed phenotype was not consistent in all the transgenic lines generated with each construct (**Figure i7 A**). From the YRT2ΔAB YAC, transgenic lines could be obtained carrying both a darker and a lighter coat colour, a result difficult to interpret but suggesting that additional regulatory elements might lay in the locus, or that these mutated YAC transgenes became sensitive to chromosomal position effects. Nonetheless, lines could be generated that show pigmentation levels similar to that of the YRT2 lines, indicating that correct expression can be achieved when suboptimal constructs are integrated in a favourable genomic locus (Giraldo and Montoliu, 2002 and Giraldo, PhD thesis, 1999).

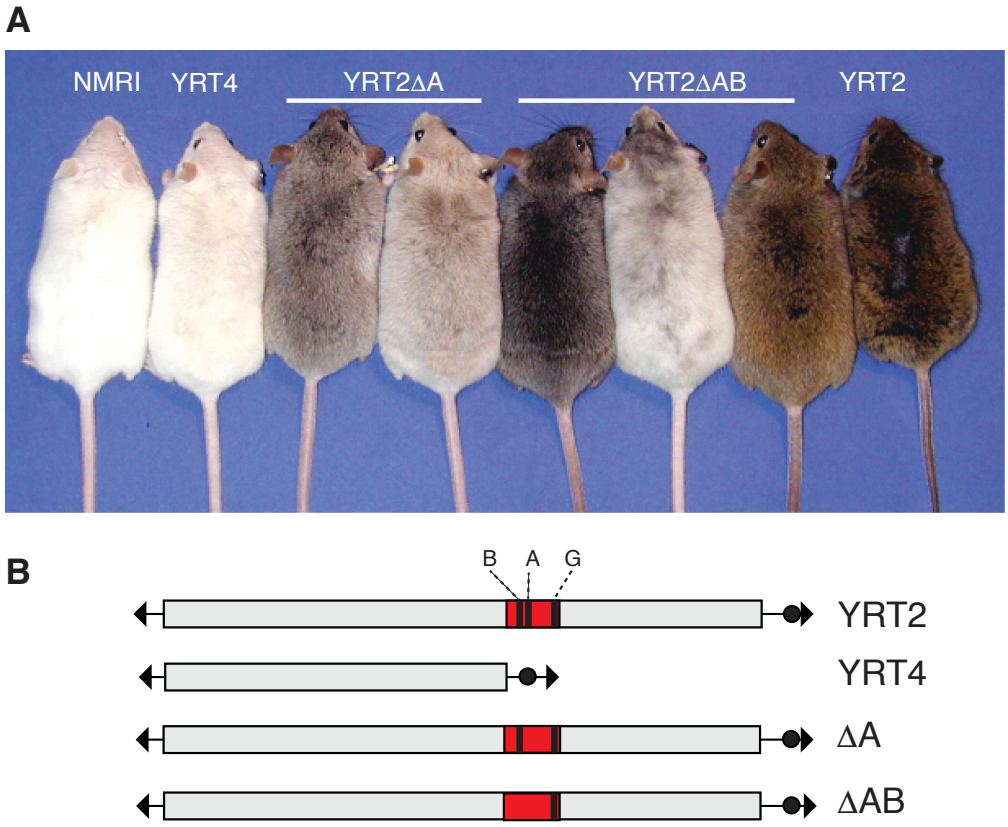


Figure i7 Tyr-YAC transgenic mice lacking the A and AB boxes. **A** NMRI albino and YRT4 mice are shown as control. Two individuals from two independent transgenic lines generated using the YRT2ΔA resulted in distinct levels of pigmentation. Three independent lines generated using the YRT2ΔAB display coat color alterations that are highly variable between lines. A YRT2 individual is shown as control. **B** Scheme of the transgenes used in A. Telomers are indicated as arrowheads, centromers as circles.

Although milestone results have been obtained with the inactivation of *cis*-acting element within large transgenes, this approach suffers of several limitations that might jeopardize the result obtained. Chromosomal-type transgenes such as YAC and BAC are considerably larger than plasmid-based transgene and normally result in optimal expression levels (Giraldo and Montoliu, 2001; Van Keuren et al. 2009). Nevertheless, BACs and YACs still behave as a transgene: they integrate randomly in the genome, sometimes also in multiple copies and in a non-controllable fashion, and can undergo rearrangements. Transgene integrity, copy number and the integration site should be determined for each newly generated transgenic line, which is technically demanding and time-consuming and often not possible. Certain variability can be also observed between lines generated with the same construct (**Figure i7A**; Giraldo, 2002 PhD thesis), depending on the integration site and copy number. This observation represents a *caveat* to be kept in mind when comparing transgenic lines

generated using variants of the same large transgene, as done for the YRT2-derived constructs.

From a merely technical point of view, large constructs that accommodate multi-gene large chromatin domains (such as gene clusters) might be hard to obtain and difficult to handle, modify and deliver to the mouse embryo.

In addition, the disruption of boundary elements – such as the mouse *Tyr* LCR - will likely affect the regulation of close-by genes (Volpi et al. 2012), but also the organization of the chromatin of the entire topological domain and, in some cases, also the expression of genes far away from the inactivated element. For these reasons it is desirable to inactivate *cis*-acing elements in their endogenous location and not on a large transgene, in order to minimize the number of experimental variables the must be taken into account. However, it must be noted that the used of genomic-type transgenes has greatly facilitated the analysis of *cis*-acting elements *in vivo* (Giraldo and Montoliu, 2001).

1.2.2 Inactivation of *cis*-acting elements using RNA interference

RNA interference (RNAi) technology (Fire et al. 1998) is routinely employed to achieve transient partial depletion of a given protein, known as gene knock-down (KD). Gene KD by RNAi is obtained by various cellular mechanisms including endonucleolytic cleavage of the gene transcript mRNA and inhibition of translation initiation. In addition, RNAi is capable of inducing *de novo* methylation of the targeted DNA region (Verdel et al. 2004; Buhler et al. 2006). Targeted, siRNA-mediated chromatin remodelling was described in mammalian cells (Morris et al. 2004; Kawasaki et al. 2005; Kim et al. 2006) and used to inactivate DNA regulatory elements in mouse and human cells (Gonzalez et al. 2006). The delivery of a pool of short interfering (si)RNAs targeting the INK4/ARF regulatory domain ($RD^{INK4/ARF}$) triggered the increase of the repressive histone mark H3K9me3 at the targeted DNA sequence as well as a decrease in the transcript and protein levels of the three genes regulated by the $RD^{INK4/ARF}$(Gonzalez et al. 2006).

In our laboratory, RNAi was used to target the key sequences found within the *Tyr* LCR, in the tyrosinase-expressing mouse melanoma B16 F1 cells (**Figure i8**). Using the MicroRNA-adapted retroviral vector MSCV/LTRmiR30-PIG (Dickins

et al. 2005), siRNA were designed targeting the AB and G boxes (Giraldo et al. 2003) along with control siRNA targeting the *Tyr* promoter and first exon.

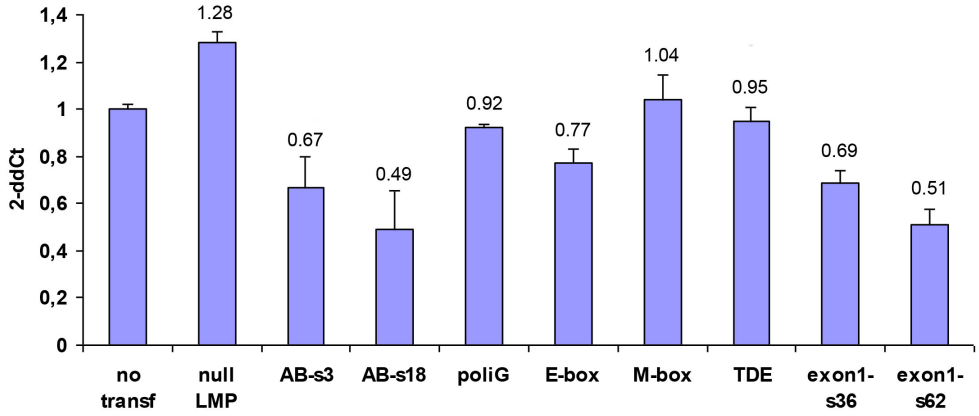


Figure i8 Changes in Tyr expression induced by RNAi. The LCR was targeted using RNAi vectors specific for the AB boxes (AB-s3 and AB-s18) and the G box (poliG). As positive controls, RNAi constructs targeting the three MITF-binding site in the Tyr proximal regulatory elements were used (E-box, M-box and TDE). Similarly, sequences of the Tyr first exon were used as positive control (exon1-s36 and s-62). Adapted from Moltó and Montoliu, unpublished.

Although a reduction in *Tyr* expression was achieved upon delivery of RNAi vectors specific for the AB boxes and for two sequences in *Tyr* exon 1, targeting basic and fundamental proximal regulatory elements in the *Tyr* promoter failed to perturb significantly gene expression. In addition, due to difficulties in the genetic manipulation in B16 cells, stable expression of the RNAi vectors could not be achieved and key experiments – such as the analysis of the deposition of repressive histone marks after RNAi – could not be performed.

1.2.3 Chromosomal inactivation using gene targeting in mouse ES cells

Classical gene targeting by homologous recombination in embryonic stem cells represents the golden standard technique to produce a wide variety of custom site-specific mutations in mice over the last 30 years (Capecchi 2005). These include subtle point mutation, constitutive and conditional gene inactivation (knock-out, KO and cKO), chromosomal deletion as well as targeted integration (knock-in, KI). Generation of targeted mutant mice is a routine but complex process based on accurate design of targeting vectors, standardized ES cell culture and transfer of selected, recombinant ES cell clones into blastocysts, that will contribute to development and hopefully colonize the mouse germ line (**Figure i9**).

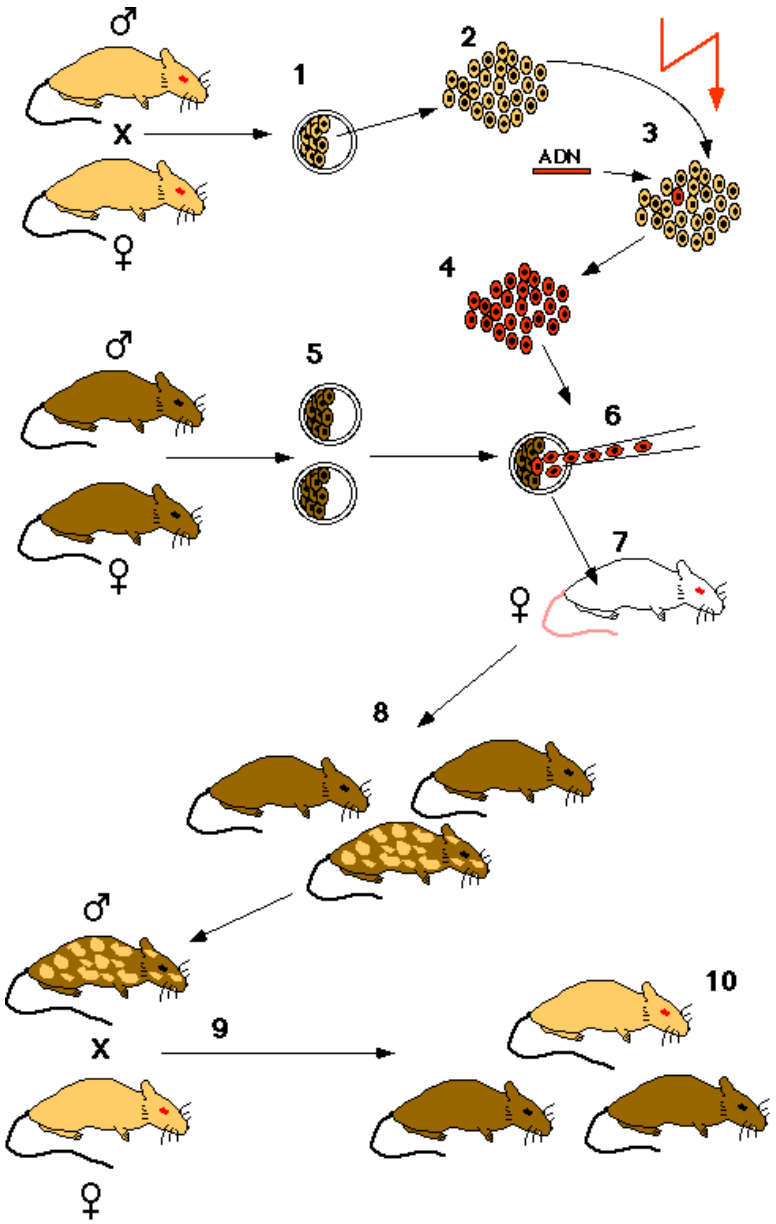


Figure i9 Gene targeting in mouse ES cells. **1** Blastocysts are obtained from a donor mouse strain and **2** ES cell lines are established. **3** DNA constructs targeting a given locus are delivered to the ES cells and **4** recombinant clones carrying the desired genetic alteration are selected and expanded. **5** Blastocysts are obtained from the recipient mouse strain, with a different coat colour genetic background. **6** A number of ES cells from the selected clones are injected into recipient blastocysts and **7** transferred into pseudopregnant females. **8** The obtained animals are chimeras. Coat colour contribution indicates the percentage of cells derived from recipient blastocyst and injected ES cells. **9** Breeding selected chimeras with individual from the donor mouse strain is necessary to check for germline transmission of the mutation in a isogenic background. **10** The resulting, targeted mutant animal will have the same coat color as that of the original donor mouse in 1. Adapted from Montoliu, 2000.

The vectors used for gene targeting contain two large stretches of DNA that are homologous with the region to be targeted. For the targeting event to happen efficiently, these homology arms must be obtained from isogenic DNA (te Reile et al. 1992; Andreasson et al. 2013) and should not contain repetitive DNA. Upon homologous recombination, with a frequency ranging between 10^{-5} to 10^{-6} (Reid et al. 1991), the genomic sequence located in between the homology arms will be

replaced by the sequence flanked by the same homology arms in the targeting vector. Positive and negative selection cassettes are included inside and outside the homology arms, respectively, in order to select those rare cells where the correct recombination event had occurred (**Figure i10**). Chromosomal inactivation of a non-coding region can be obtained by mutating a known, relevant transcription factor binding site, or – more frequently – by a large chromosomal deletion including all putative protein binding sites. Obtaining a chromosomal deletion requires a targeting vector carrying homology arms that flank the sequence to be deleted. Alternatively, a positive selection cassette can be used to replace the region to be deleted. If wisely flanked by loxP sites, the selection cassette can be removed using Cre recombinase, either *in vitro* (Ryder et al. 2013) or *in vivo* using ubiquitous Cre-driver mouse lines (Nagy, 2000).

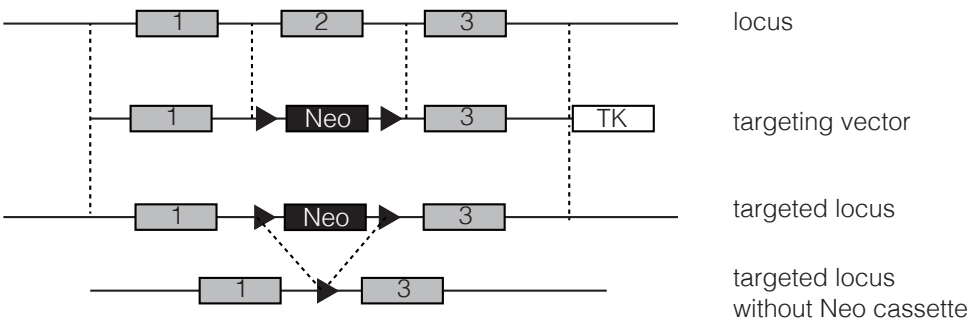


Figure i10: Classical strategy for targeted chromosomal deletion. A targeting vector containing large homology arms flanking the DNA sequence to be deleted is delivered to mouse ES cells. Positive selection can be applied thanks to a Neomycin resistance gene cloned inside the homology arms. Randon integration of the targeting vector can be prevented by using a negative selection cassette, cloned outside the homology arms (in this case, a Thymidine kinase cassette), Cre-mediated recombination is then used to remove the positive selection cassette that was previously flanked with loxP sites.

The inactivation of the mouse β-globin LCR, performed by the group of Mark Groudine (Fred Hutchinson Cancer Research Center, Seattle WA, USA) is one of the attempts to study regulatory element by inactivating them *in vivo* (Bender et al. 2000). The human and mouse β-globin genes are ordered in a linear array that reflects their order of expression. The β-globin locus is embedded in a large array of olfactory receptor genes, expressed in sensory neurons of the nasal epithelium (Buck and Axel 1991). DNAse HS sites were found at both 5' and 3' extremities of the β-globin genes (Bulger et al. 1999) (**Figure i11**). The series of HS sites at 5' acts as a strong erythroid-specific positive regulatory element and therefore defined as LCR (Grosveld et al. 1993). By several round of gene targeting, using a suitable combination of selection markers, the authors could

delete a 24 kb region containing the LCR sequence. In the absence of the LCR sequence, the expression of the β-like globin genes is reduced to the 2-8% of the normal expression level. Interestingly, even if at lower levels, all the β-globin genes can be detected at the proper tissue and developmental stage, even in the absence of the LCR. Analysis of the chromatin state in mice carrying the LCR deletion indicated that the LCR does not contribute to the establishment and to the maintenance of the open chromatin state of the locus. Rather, the LCR is necessary to reach the physiogical expression levels required to fulfill the biological function underneath.

Figure i11 Diagram of the mouse β globin locus. The β globin gene cluster is embedded in a olfactory receptor genes array. Several DNAse I hypersensitivity sites were identified at both sides of the β globin cluster, indicated as black arrowes. The β globin LCR, conserved in the human genome, is found at 5'. DNAse I hypersensitivity sites are. Genes position and CTCF/Rad21 occupancy were extracted from the UCSC Genome Browser. Adapted from Burger et al. 1999.

Using a similar strategy, Volpi and colleagues deleted 8 kb of an insulator element at the 3' end of the mouse *Igh* locus. Upon deletion, no histological or haematological alterations were observed in transgenic animals compared with controls. Interestingly, 65 kb downstream the deletion, a non-*Igh* gene, *Hole,* was found to be overexpressed during B-cell stimulation (Volpi et al. 2012). Notably, several control experiments were required to ascertain that the phenotype observed was due to the deletion of the *cis*-acting sequence and not to genetic differences between the ES cell genetic background, the recipient background and that of the cohort of animals used as control.

These works and other (Danielian et al. 1997; Yanagisawa et al. 2003; Visel et al. 2010; Shim et al. 2012) demonstrated that gene targeting is ES cells can be used produce *cis*-ruptions in the natural chromosomal surroundings of many loci, in mice. Nevertheless, the generation of such animal models is technically demanding, complex and time consuming. Notably, a key aspect for successful gene targeting is the choice of optimal homology arms to be included in the targeting vector. Homology arms length may range from 500 bp to 2 kb (or

larger) and should be obtained from isogenic DNA, be free of repeated elements and match perfectly (and uniquely) with the sequence to be targeted. Due to the presence of large stretches of repetitive DNA in the mouse *Tyr* locus (**Figure i12**) the design of an optimal targeting vector, devoid of DNA repeats, is complicated. In fact, in our laboratory, a previous attempt to obtain the *Tyr* 5' deletion using classical gene targeting approaches failed (Tovar and Montoliu, unpublished).

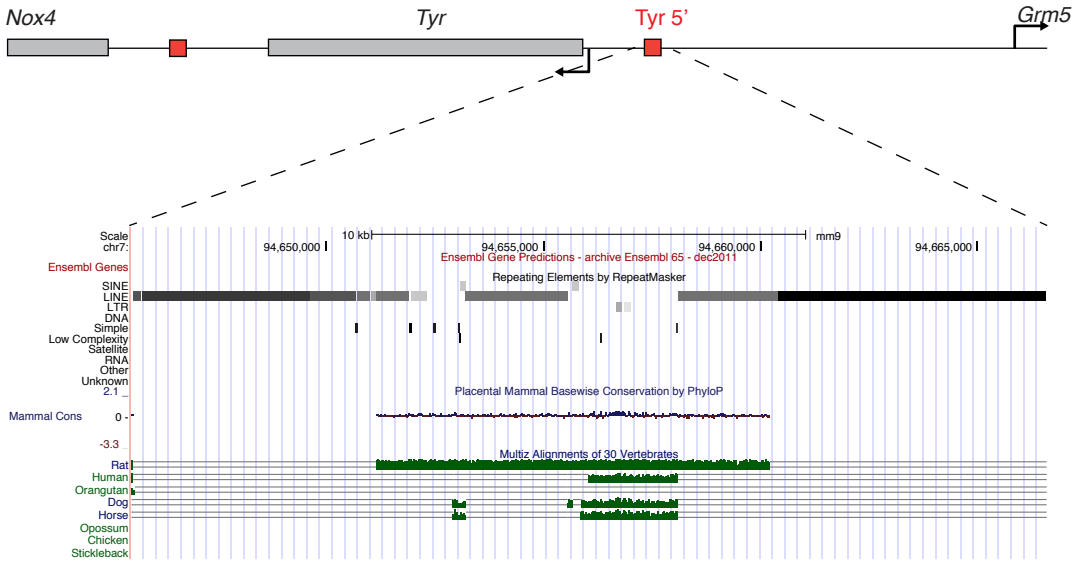


Figure i12: Repetitive sequences flanking the mouse *Tyr* 5' boundary. Schematics illustrating sequence conservation and content in repeat elements, obtained from RepeatMasker and displayed using UCSC Genome Browser. Repetitive DNA sequences are shown as black and grey boxes. The *Tyr* 5' boundary is surrounded by LINE-1 repeats. The LINE-1 repeat at 3', described in Giraldo 2003, is methylated in both B16 mouse melanoma and L929 mouse fibroblast cells.

Using the targeting strategy indicated in **Figure i10**, CK35 mouse embryonic stem cells (Kress et al. 1998) were electroporated with a target vector. More than 400 Neomycin-resistant colonies were screened. Eventually, by Southern Blot analyses two ES cells colonies were found displaying the expected restriction pattern (**Figure i13**). After microinjection into mouse blastocysts, a total of 4 chimeras were obtained that were able to transmit the targeted allele to the progeny. Nevertheless, upon crossing with an albino Cre-driver mouse, no alteration in pigmentation was observed. Careful molecular analyses indicated that the targeting happened on at a different chromosomal location, highly homologous with the region found 5' in the *Tyr* locus, but – unfortunately – not on the actual *Tyr* locus on chromosome 7, as shown by FISH analyses (**Figure i14**).

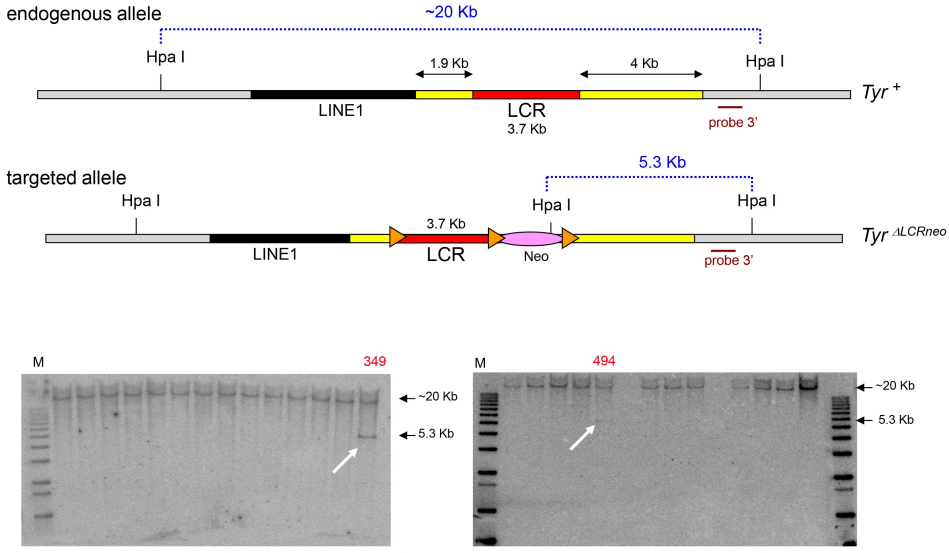


Figure i13 Targeting the *Tyr* 5' boundary in mouse ES cells. Homology arms are indicated in yellow. Using the probe 3' on HpaI-digested genomic DNA, a 20 kb band is expected for wild-type clones and a second, 5.3kb large band is expected for targeted clones, as shown for clones 349 and 494. Adapted from Tovar and Montoliu, unpublished.

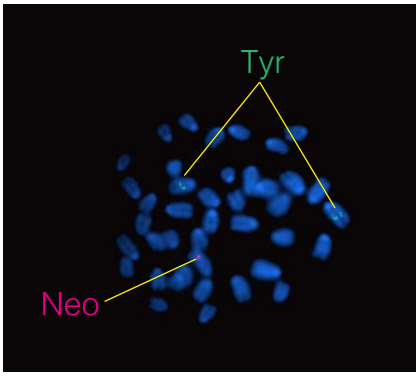


Figure i14 FISH analysis. Due to extreme repeats in the *Tyr* locus, targeting to incorrect positions is frequent event. A FISH image is shown. No colocalization is detected between a *Tyr* probe (in green) and a *Neo* probe (in red). Adapted from Tovar, Nadal and Montoliu, unpublished.

1.2.4 Chromosomal inactivation using targeted nucleases

As an alternative to RNAi or gene targeting in mouse ES cells, chromosomal deletions can be generated in mammalian cells and embryos using targeted nucleases such as Zinc Fingers (ZFNs), Transcription Activators-Like Effector Nucleases (TALENs) and, recently, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) (Seruggia and Montoliu, 2014). Targeted nucleases are engineered proteins able to recognise – by different mechanisms – a short unique DNA sequence and induce a double strand break (DSB) at the targeted sequence. The sequence constraints associated with the use of targeted

nucleases are minimal; as much as twenty to sixty base pairs are necessary to define a targetable site in a vertebrate genome. In contrast, for classical gene targeting, two 500 bp-2 kb large unique sequences must be identified at both sides of the target site, a condition that can't be fulfilled in many loci and particularly this is true for most non-coding region, as illustrated in the previous paragraph with the *Tyr* locus.

Using targeted nucleases, genetic modifications can be obtained *in vivo* by microinjecting the nuclease (as DNA, mRNA or protein) into mouse fertilized eggs. A wide variety of genetic modifications can be obtained using individual targeted nucleases, multiple nucleases or in combination with donor DNA molecules (**Figure i15**). Nuclease-induced DSBs can be repaired by homology directed repair (HDR) or by the error-prone non-homologous end-joining pathway (NHEJ, Barnes 2001). In the presence of a donor plasmid or a single stranded oligonucleotide, HDR can lead to the targeted introduction of a transgene (KI) or to target subtle base changes. In the absence of a DNA repair template, NHEJ-mediated DNA repair will result in small insertions and deletions (*indels*) that might cause the disruption of the targeted gene.

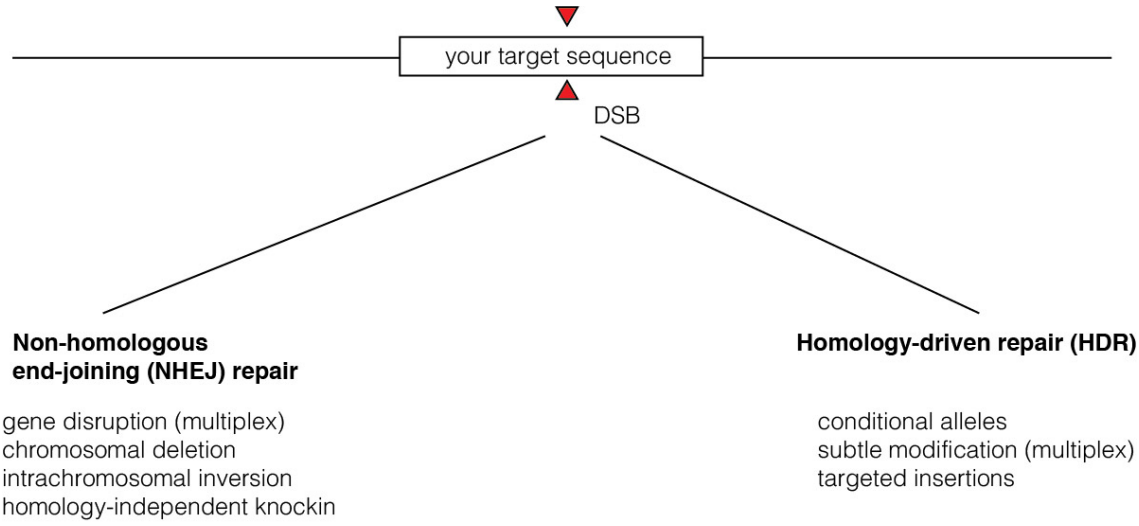


Figure i15 Editing possibilities using targeted nucleases. After a targeted double strand break is produced, two alternative DNA-repair pathways can be exploited to obtain a variety of genetic manipulations. Gene Knock Out (KO), chromosomal deletions and inversions can be obtained in the absence of a DNA donor template. Conditional Knok Out (cKO), subtle point mutations and targeted insertions can be achieved through the homology-driven repair pathway in the presence of a DNA donor template. Adapted from Seruggia and Montoliu, 2014.

When two DSBs are generated on distal positions on the same DNA molecule, the intervening DNA can be excised (or inverted) and the two DNA

ends joined by NHEJ or by micro-homology mediated end joining (MMEJ, Lieber et al. 2003) (**Figure i16 A**).

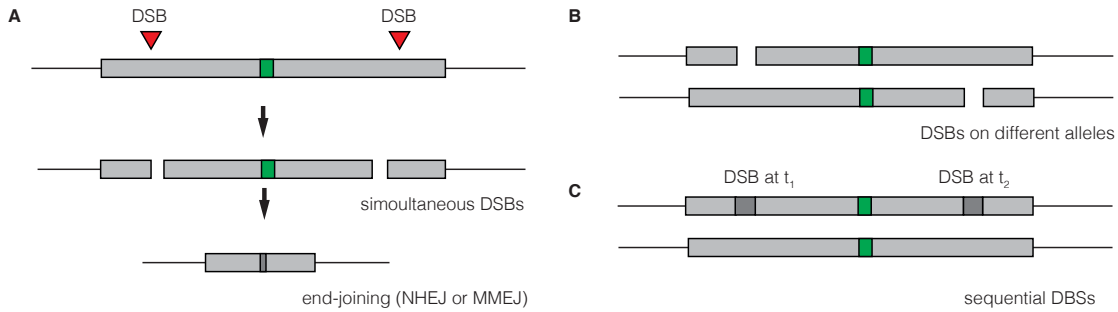


Figure i16 Strategy to produce chromosomal deletions using nucleases **A** Two adjacent DSBs result in deletion of the intervening DNA. **B** Only DSBs occurring on the same DNA molecule, in *cis*, can yield a deletion. **C** Two DSBs must be produced within a short time frame. Namely, the second DSB must happen before the first one is repaired.

This approach can be challenging, considering that both nucleases must cleave the respective target sites on the same allele, not on different alleles (**Figure i16 B**). Also, the two DSB events must happen simoultaneously or within a short interval of time, namely the second DSB before the first one has been repaired (**Figure i16 C**).

In the next paragraphs, each of the three main targeted nucleases family (ZFN, TALEN and CRISPR-Cas9) will be introduced, along with the attempts to produce targeted, chromosomal deletions using each tool, reported to date.

1.2.4.1 Zinc Finger Nucleases (ZFNs)

Zinc Finger Nucleases (from now on, ZFNs) are the first class of targeted nucleases that were generated and therefore the most characterized to date. ZFNs are composed of a custom DNA-binding motif coupled with the catalytic domain of the *Fok*I endonuclease (Bibikova et al. 2002; Urnov et al. 2010) (**Figure i17**). The DNA-binding motif used in ZFNs is a Cys_2-His_2 class of zinc-finger domain, which is the most common DNA-binding motif in eukaryotes (Rubin et al. 2000) and the second most frequently encoded type of protein domain in the human genome. Individual fingers contain an average of 30 amino acids and recognise 3 base pairs each. Arrays of zinc fingers can be assembled using artificial peptide linkers (Liu et al. 1997; Moore et al. 2001) and can be designed to recognise unique targets in

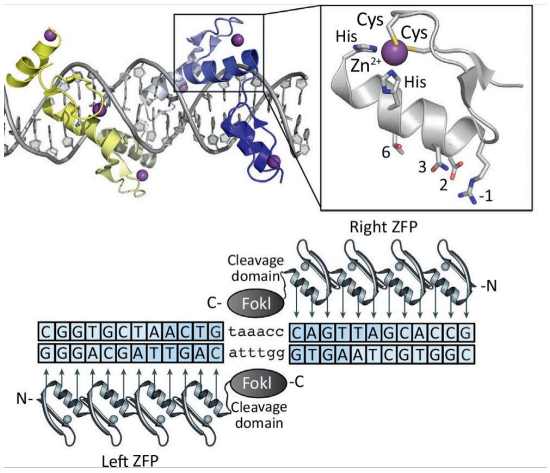


Figure i17 Structure of Zinc Finger Nucleases. Designed zinc-finger protein in complex with target DNA (grey). Each zinc-finger consists of approximately 30 amino acids in a ββα arrangement (inset). Surface residues (-1, 2, 3 and 6) that contact DNA are shown as sticks. Each zinc- finger domain contacts 3 or 4 bp in the major groove of DNA. The side chains of the conserved Cys and His residues are depicted as sticks in complex with a Zn^{2+} ion (purple). A cartoon of a ZFN dimer bound to DNA. ZFN target sites consist of two zinc-finger binding sites separated by a 5–7-bp spacer sequence recognized by the *Fok*I cleavage domain. Zinc-finger proteins can be designed to recognize unique 'left' and 'right' half-sites. Adapted from Gaj et al. 2013

complex genomes, such as rat. (Geurts et al. 2009). Designing optimal zinc-finger arrays is a complex task that can be achieved by modular assembly of pre-selected zinc-fingers recognising individual nucleotide triplets (Beerli et al. 2002; Bhakta et al. 2013). However, the affinity and specificity of each zinc finger towards the corresponding triplet may vary because of context-dependant interactions between neighbouring fingers. This observation led to the development of selection-based approaches to empirically choose the best array from large, pre-existing arrays libraries. The principal algorithm of ZFNs design is proprietary of Sigma Aldrich/Sangamo (St. Louis, MO, USA). Alternately, non-proprietary algorithms are available that can be used for ZFNs design and assembly (Hermann et al. 2012).

The *Fok*I nuclease domain used in ZFNs has been extensively engineered in order to increase the specificity and the mutagenic activity of ZFNs and reducing its toxicity. The development of heterodimeric mutants allowed two ZFN monomers to bind the target sequence in order to reconstitute an active FokI protein (Guo et al. 2010; Doyon et al. 2011). Furthermore, directed evolution was used to generate a hyperactive FokI variant, 15-fold more active than the original protein (Guo et al. 2010).

Chromosomal deletions using two ZFN pairs targeting distal sequences have been reported *in vitro*, but not yet *in vivo*. Liu and colleagues produced a 240 bp deletion in the DHFR gene in CHO cells, although with low efficiency (Liu et al. 2010). Similarly, deletions were generated in human cells using combinations of

ZFN pairs (Sollu et al. 2010). Lee and colleagues reported that deletions in the range of tens, hundreds and thousands of kilobases could be produced and detected in human cells (Lee et al. 2010). Notably, ZFNs producing non-compatible overhangs generated chromosomal deletions, indicating that cellular DNA end-processing event occur after nuclease-mediated DNA cleavage and that this cellular process is involved in the sealing of the two DNA extremities. In transgenic plants a transgene flanked by four copies of the ZFN target sequence could be efficiently deleted by breeding with a second transgenic plant stably expressing a single corresponding ZFN (Petolino et al. 2010).

1.2.4.2 Transcription Activators-Like Effector Nucleases (TALENs)

Transcription Activators-Like Effector Nucleases (TALENs), represent the second generation of targeted nucleases (Joung and Sander, 2013). TALENs are, similarly to ZFNs, composed of an artificial DNA-binding domain coupled to a FokI nuclease domain (**Figure i18**). The mechanism of action is identical to that of ZFNs: two TALEN monomers recognise and bind the target DNA sequence. Upon DNA-binding, two inactive FokI monomers form an active dimer that mediates a targeted DSB at the desired locus.

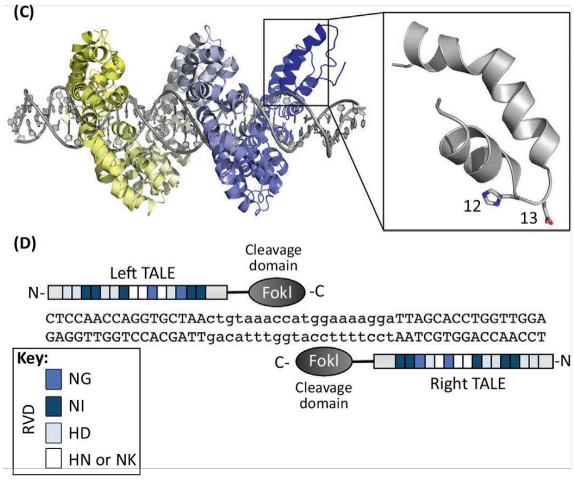


Figure i18 Structure of TALENs. TALE protein in complex with target DNA (grey). Individual TALE repeats contain 33–35 amino acids that recognize a single base pair via two hypervariable residues (repeat-variable diresidues; RVDs) (shown as sticks) (inset). A cartoon of a TALE nuclease (TALEN) dimer bound to DNA is shown. TALEN target sites consist of two TALE binding sites separated by a spacer sequence of varying length (12–20 bp). TALEs can be designed to recognize unique left and right half-sites. RVD compositions are indicated. Adapted from Gaj et al. 2013

A TALEN contains the same *Fok*I domain previously used for ZFNs. On the contrary, the DNA-binding domain contained in TALENs is radically different. It is

derived from a series of secreted proteins found in *Xanthomonas*, called Transcription Activator-Like Effectors (TALEs) that these pathogens use to alter gene expression in the hosting

plant cells, upon infection (Boch et al. 2010). The sequences of several TALE proteins were aligned and found to be composed of arrays of highly conserved 33-35 amino acids repeats carrying two central hypervariable residues at positions 12 and 13 (**Figure i19**).

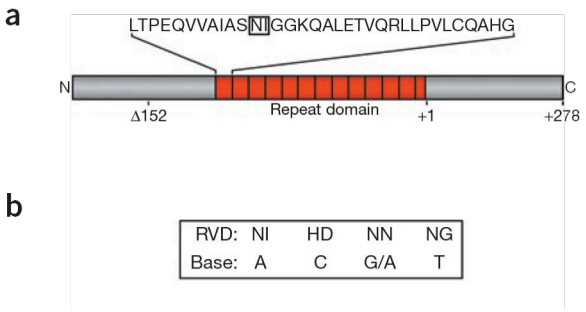


Figure i19 Structure and DNA-binding specificity of TALE proteins. **A** Sketch of a TALE from Xanthomonas. Red rectangles indicate the array of tandem repeats. The sequences of the first array is indicated, with the amino acids at position 12 and 13 boxed. **B** The protein-DNA code at the basis of the TALE DNA-binding specificity. Adapted from Miller et al. 2010.

Following a specific code (Moscou et al. 2009), repeats carrying different amino acids at position 12-13 were found to bind a single DNA base (Boch et al. 2009). Custom arrays of 15-30 repeats can be assembled seamlessly in order to design DNA-binding proteins with the desired specificities. The transparency of the protein-DNA code, the public distribution of reagents from the original laboratories (Cermak et al. 2011) and the development of efficient cloning techniques (Engler et al. 2009, and sections 3.1.4 and 3.1.10), determined the rapid success of TALEN in the genome-engineering field, as compared with ZFNs.

Large chromosomal deletions could be obtained in cultured cells by co-transfection of two TALEN pairs targeting distal sites on the same chromosome (Ma et al. 2014). Few reports have described the production of large deletion using TALENs in mice. By injecting the four RNA molecules necessary to express two TALEN pairs, Svodoba and colleagues generated a transgenic mouse carrying a targeted deletion of an intronic retrotransposon, that is responsible of driving the expression of a Dicer isoform specifically expessed in oocytes (Dicer^{O}, Flemr et al. 2013). Targeted deletions are much more frequently reported in zebrafish (Ma et al. 2013; Lim et al. 2013; Liu et al. 2013; Xiao et al. 2013), an organism that tolerates the injection of higher amount of nucleic acids, as compared to mice. Ekker and colleagues reported the efficient production of

deletion in the range of 16-18 kb by microinjection of two TALENs pairs in zebrafish embryos (Ma et al, 2013).

1.2.4.3 CRISPRs

CRISPR/Cas9 is the latest piece of genome editing to be included, very recently, in the genome engineering toolbox. The basis or this system is derived from the world of microbiology, similarly to TALENs: CRISPRs (clustered regularly interspaced short palindromic repeats) sequences and Cas (CRISPR-associated) proteins are the two elements of an ancient prokaryotic adaptive defensive system conserved in archaeal and other bacterial genomes (Jansen et al. 2002; Makarova et al. 2011; Jinek et al. 2012) (**Figure i20**). CRISPR sequences are accumulated in a specific locus at every infection event and represent the memory of the system; they form a repository of short, directly repeating nucleotide sequences that alternate with small unique DNA fragments (Ishino et al. 1987; Bolotin et al. 2005). Cas proteins process CRISPR sequences into small RNAs (Haurwitz et al. 2010) that are used to recognise sequence of infectious agents at a successive event of infection (Haft et al. 2005). Finally, Cas proteins cleave the infectious DNA molecules that match the CRISPR-derived RNA (Maraffini and Sontheimer 2008; Garneau et al. 2010; Gasiunas et al. 2012).

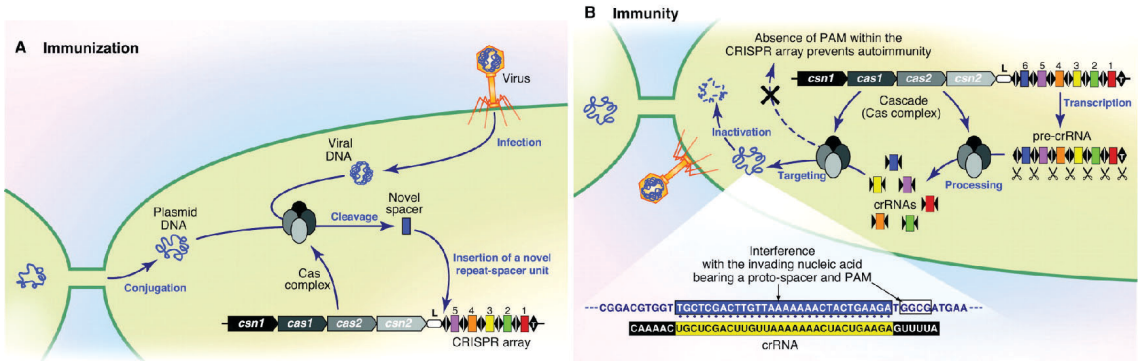


Figure i20 CRISPR/Cas restriction system. **A** Upon infection, the exogenous genetic material of the invader is processed by a Cas complex. Small fragments of the exogenous DNA are integrated in the CRISPR locus. **B** Small crRNAs are generated from the CRISPR locus. When a DNA matching with one crRNA is found, a second Cas complex cleave the DNA of the invader. Adapted from Horvath and Barrangou, 2010

The two main RNA molecules required for CRISPR to work, namely the crRNA (CRISPR RNA) and the tracrRNA (trans-activating crRNA) were fused into a single sequence (the sgRNA, small guide RNA) to facilitate the translation of the CRISPR system to the genome editing field (Jinek et al. 2012). Thus, the sgRNA contains an 80 bp hairpin RNA structure that resembles the tracrRNA, connected to a 20 bp sequence homologous to the target sequence (the crRNA). Among several elements of the Cas proteins (Chylinski et al. 2013), Cas9 is sufficient to complex with the CRISPR RNA and the DNA target (Nishimasu et al. 2014; Jinek et al. 2014; Anders et al. 2014) and to cleave both target DNA strands upon Watson&Crick homologous base pair matching between the target DNA and the sgRNA. Few sequence constraints are required for CRISPR/Cas9 genome editing. The PAM (Protospacer adjacent motif) NGG must be found 3' to the target sequence, that should be as long as 20 bp. An example is given for the mouse *Tyr* exon 1 (**Figure i21 A**). Several Cas9 and sgRNA expression vectors are available (Ran et al. 2013). Typically, CRISPR/Cas9 induces a DSB 3 bp upstream the PAM motif (**Figure i21 B**). Alternatively, a Cas9-nickase variant can be used together with two sgRNA vectors. The Cas9-nickase carries a poin mutation (D10A) that inactivates one of the two nuclease domains of the protein. As a consequence, the nickase will only impose non-mutagenic single strand breaks (nicks, **Figure i21 C**). When two nicks are produced close-by with the proper distance and orientation (**Figure i21 D**), similar signals of DNA break are produced that are able to trigger *indel* formation at the targeted site (Ran et al. 2013; Shen et al. 2014).

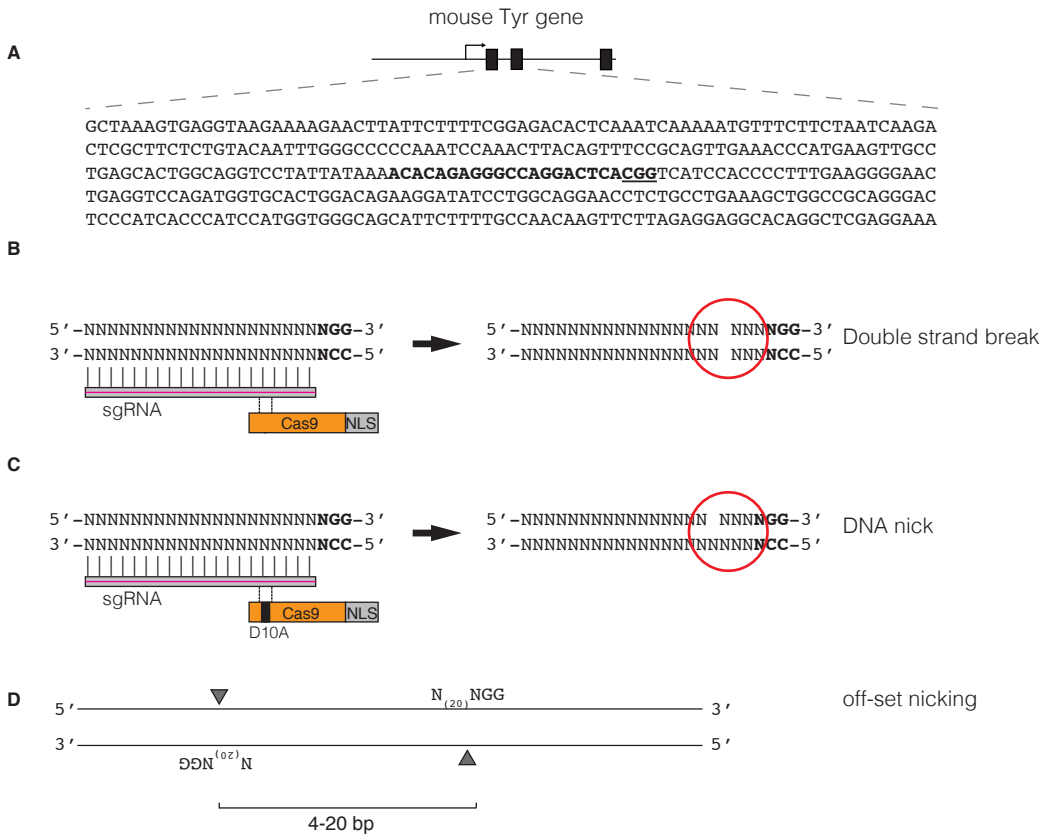


Figure i21 Mechanism of CRISPR/Cas9 genome editing. **A** Example of a potential CRISPR target site in the mouse Tyr exon 1. A N_{20}–NGG sequence is shown in bold. The PAM motif is underlined. **B** Cas9 protein, complexed with a small guide RNA specific for a desired locus, mediates a double-strand break 3 bp upstream the PAM motif. **C** If the D10A Cas9 mutant (referred as ''nickase'') is used, a single strand break (or nick) is produced. **D** In order to obtain a mutagenic DBS, two nicks must be directed using two guide RNAs located as close as 4 bp and as far as 20 bp with each other, and designed being a first sgRNA complementary to the forward strand of the target, followed by a guide RNA complementary to the minus strand, 3' downstream. Adapted from Seruggia and Montoliu 2014.

The first transgenic animal generated using CRISPR/Cas9 system was reported in the first half of 2013. Huang and colleagues used CRISPR/Cas9 to disrupt an EGFP transgene in mouse and in zebrafish embyos (Shen et al. 2013). Using this system, multiple genes can be disrupted simoultaneously by co-injection of several sgRNA molecules with Cas9 mRNA. Wang and colleagues targeted up to five genes in one single experiment (Wang et al. 2013). Such experiments illustrate the great potential of CRISPR/Cas9 system, although the genetics of the obtained animal models can be difficult to handle in practice, with different alleles segregating independently. Achieveng the same result using TALEN or ZFN is technically demanding, as the total amount of RNA involved would easily reach toxic concentrations before each individual nuclease would reach its working concentration. In fact, multiple genome editing using TALENs or ZFNs has not

been reported so far. Multiplex CRISPR/Cas9 genome editing is mostly useful to target two genes on the same chromosome in mouse (Zhou et al. 2013) and rat (Li et al. 2013), that would otherwise require – using standand techniques - several breeding rounds of two independent KO mouse lines, or two sequential rounds of gene targeting in embryonic stem (ES) cells using compatible selection strategies.

Genetic manipulation using CRISPR/Cas9 is robust and reproducible. But still, key methodological details remain to be standardized, including correct RNA concentration (or range) and microinjection route (pronuclear-only, cytoplasmic-only, or a combination of the two). This problem was addressed on a recent report from Horii and colleages. After having compared different microinjection protocols, the authors concluded that cytoplasmic-only microinjections yield the best results (Horii et al. 2014). Nevertheless, injection of some RNA material in the pronuclei to observe pronuclear swelling will serve as positive control of the microinjection. To further expedite genetic manipulation of mice and rat using haploid genomes (Li et al. 2012), CRISPR/Cas9 were successfully used in mouse (Horii et al. 2013) and rat (Li et al. 2014) haploid ES cells.

Although already possible using the previous generation of nucleases (Meng et al. 2008; Bedell et al. 2012), zebrafish KO lines could be easily generated using CRISPRs (Jao et al. 2013).

Gross chromosomal rearrangements, including deletions and inversions, can be obtained targeting distal sequences on the same chromosome. Targeted deletions were obtained in mouse zygotes (Fujii et al. 2013) and haploid ES cells (Horii et al. 2013). In zebrafish, a 40 kb DNA region was deleted and inverted (Xiao et al. 2013). Such approach is mostly useful to study large non-coding regulatory elements that might need full deletion, rather than point inactivation, to observe the underlying phenotype, as in the case of the *Tyr* regulatory elements. With the CRISPR/9Cas system, many of the non-coding elements described in the ENCODE project (Bernstein et al. 2012) or other non-coding genes, such as lncRNAs, can be functionally validated *in vivo*.

1.3 Open questions and technical advances

The *Tyr* locus is an optimal experimental model to study the mechanisms that control gene expression. In fact, *Tyr* expression pattern is very specific, as introduced at the beginning of this section, and can be directly traced, being itself a reporter gene (Beermann et al. 1991) as well as a cell lineage marker (Gimenez et al. 2003).

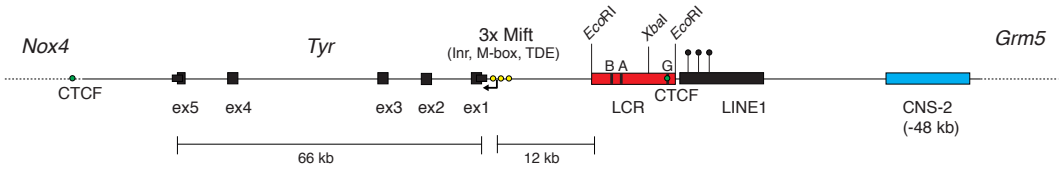


Figure i22 Known regulatory elements in the mouse *Tyr* locus. The five *Tyr* coding exons are shown as black boxes and spans a region of 66 kb. Three Mitf-binding sites identified on the *Tyr* promoter are depicted as yellow circles. The LCR, indicated as a red box, is a 3.7 kb *Eco*RI fragment located 12 kb upstream the *Tyr* promoter and includes A, B and G boxes, depicted as black bars. Two CTCF sites were previously described in our laboratory and are indicated as green circles. A LINE1 element, indicated as black box, found immediately upstream the LCR, was found to be methylated in vitro in L929 and B16 mouse cells. A second regulatory element was identified by Murisier et al. 2007, 48 kb upstream the *Tyr* promoter, and depicted as blue box.

A considerable amount of knowledge on the *Tyr* regulatory circuits has been accumulated during the past years and summarized in **Figure i22**. Nevertheless, information is still lacking regarding the chromatin structure of the locus, the relationship with close-by genes and the existence of a second chromatin boundary, at the 3' end of the locus. These questions will be addressed in the present work using Chromosome Confromation Capture (3C), a technology used to detect the interaction of distal DNA sequences, that have noy been applied yet to study the mouse *Tyr* locus. Moveover, a series of assays, developed in our laboratory and aimed to assess the enhancer-blocking activity of putative insulator sequences, will be applied to search for a candidate boundary at the 3' end of the locus.

It is well known and demonstrated that the presence or the absence of the *Tyr* LCR in *Tyr*-YAC transgenes is a key factor to achieve faithful regulation of *Tyr in vivo*. But what happens if the LCR is *surgically* removed or inactivated from the endogenous *Tyr* locus? Will the phenotype will be consistent in all lines eventually generated, or will suffer great variability, as observed when the A and B boxes were inactivated in YAC-based transgenes? The targeted nuclease systems, reviewed in this introduction, will be instrumental to obtain the necessary genetic manipulations, that couldn't be achieved to date, after several

attempts using with both standard gene targeting techniques and RNA interference experiments.

2. Objectives

The aim of this Thesis is the functional and structural description of the mouse *Tyr* locus along with the identification of the DNA regulatory element that are necessary to recapitulate the expression of *Tyr* gene. To achieve this goal, the following objectives were addressed:

1. To analyze *Tyr* genomic sequence conservation and synteny across vertebrates to identify new potential boundary elements.
2. To validate *in vitro* previously known boundary elements and to test additional newly identified boundaries, *in vitro* and *in vivo*, using the Enhancer-blocking assays developed in our laboratory in cultured cells and in zebrafish embryos.
3. To describe the physical organization of the mouse *Tyr* locus in its active and inactive conformation by Chromosome Conformation Capture (3C).
4. To generate a series of transgenic mouse lines carrying the inactivation of the *Tyr* boundary elements in the endogenous locus
5. To analyze the phenotype of the obtained transgenic mouse lines and compare it with previously generated, *Tyr* transgenic mice available in the laboratory.

3. Methods

3.1 Original vectors and cloning

3.1.1 pELuc for Enhancer Blocking Assay in HEK 293 cells

All plasmids used in the *in vitro* enhancer-blocking assay are derivative of the pLuc (4.9 kb) and pELuc (5.5 kb, **Figure m1**) vectors, originally obtained from Dr. Satoshi Watanabe (Department of Developmental Biology, National Institute of Agrobiological Sciences, Tsukuba, Japan)(Watanabe et al. 2006). The pLuc vector contains the firefly luciferase reporter gene under the control of human cytomegalovirus minimal promoter (CMV-mP). pELuc includes, in addition, the CMV enhancer (CMV-IE). The poly(A) signal of both vectors derives from the SV40 virus.

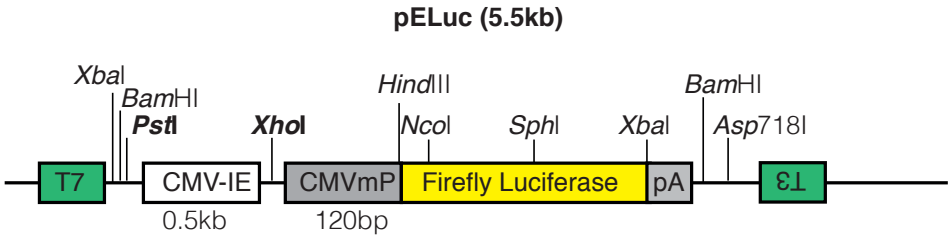


Figure m1: Map of the pELuc vector. The sequences to be tested are cloned in *Xho*I and *Pst*I position. The Asp718I site is used to linearize the vector before transfection.

The 1.2-kb chicken 5'HS4 beta-globin insulator element was used as positive control. Similarly, the internal II/III boxes from the chicken 5'HS4 beta-globin insulator element (0.2 kb), both wild type and mutated, were used as positive and negative controls, respectively (Bell et al. 1999; Chung et al. 1993; Recillas-Targa et al. 1999). In addition, pCMV-lacZ (7.2 kb) was used as a transfection control for normalization purposes (MacGregor & Caskey, 1989).

Tyr 3' (2.5 kb) *Xho*I and *Tyr* 3' (2.5 kb) *Pst*I pELuc-derivative vectors were generated by cloning a 2.5 kb sequence amplified with the primers indicated in **Appendix 9.1.1**. Similarly, a 241 bp fragment of the 2.5 kb sequence was PCR-cloned, to obtain *Tyr* 3' core *Xho*I and *Tyr* 3' core *Pst*I vectors. The *Tyr* 3' core MUT *Xho*I and *Pst*I vectors were obtained by site directed mutagenesis using the *Tyr* 3' core *Xho*I and *Tyr* 3' core PstI vectors as template. All vectors produced were confirmed by restriction digest and sequencing. Cloning procedure is described in sections 3.1.9 and 3.1.10

3.1.2 p48RCAR for Enhancer-blocking Assay in zebrafish

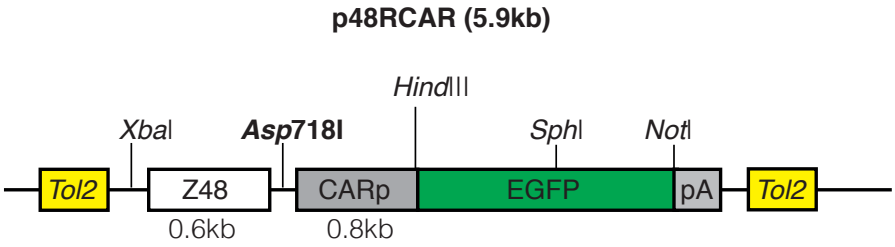


Figure m2: map of the p48RCAR vector. The putative insulator is cloned between a midbrain enhancer (Z48) and a somite promoter (CAR) using the *Asp*718I site. The enhancer/promoter/EGFP cassette is flanked by inverted *Tol2* inverted repeats for transposase-based transgenesis.

The vectors used for the enhancer-blocking assay in zebrafish are based on the p48RCAR vector (Bessa et al. 2009) (**Figure m2**). The *Xenopus laevis* cardiac actin promoter (CAR, Mohun et al. 1986) and a midbrain enhancer (Z48 or Z54390) from *Danio rerio* (de la Calle-Mustienes et al. 2005) drive EGFP expression to the somites and midbrain, respectively. *Tol2* sites for transposon-based transgenesis flank the expression cassette. The *Tyr* 3' core and its mutated counterpart were released from the *Tyr* 3' core and *Tyr* 3' core MUT *Pst*I vectors and blunt-cloned in the *Asp*718I site of p48RCAR. Cloning procedure is described at section 1.1.9

3.1.3 Bacterial Artificial Chromosomes used in this study

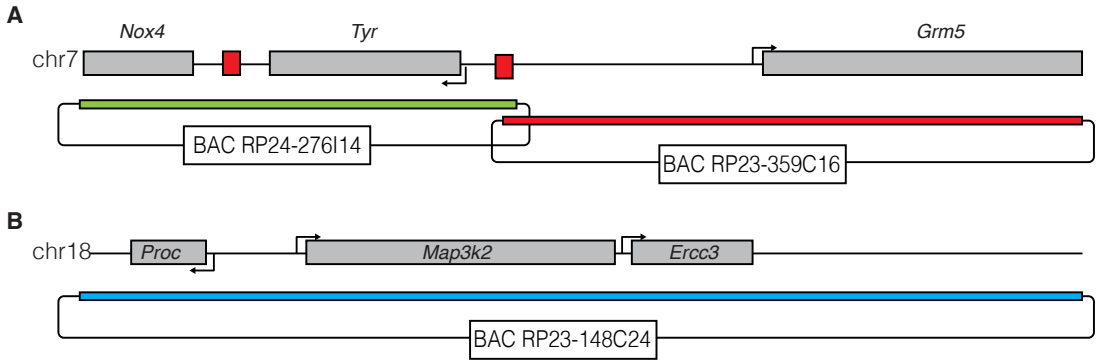


Figure m3 BACs used in this study. **A** BACs carrying the sequence of the mouse *Tyr* locus and **B** of the mouse *Ercc3* locus. All BACs are based on the pTARBAC backbone and pBACe3.6 (Osoegawa et al. 2000).

Bacterial Artificial Chromosomes (BACs) were obtained from CHORI (https://bacpac.chori.org/) and validated by end-sequencing. The presence of the sequences of interest was furthermore verified by PCR. For the Chromosome Conformation Capture study we used BAC RP24-276I14, RP23-359C16 and

RP23-148C24. The genomic content of these BACs is illustrated in **Figure m3**. BAC DNA was purified using Large-Construct Kit (QIAGEN) as indicated by the manufacturer, and stored at +4ºC.

3.1.4 TALEN Golden gate vectors

Custom TALENs were assembled by Golden Gate cloning (a procedure explained extensively in sections 3.1.9 and 3.1.10) using the Golden Gate TALEN kit (Addgene TALEN Kit #1000000024 and Cermak et al. 2011) from Dr. Daniel Voytas' laboratory. The kit contains three series of Golden Gate-compatible vectors including:

-60 *Bsa*I-based plasmids (pHD1-10; pNG1-10; pNI1-10; pNN1-10; pNK1-10; pNH1-10) containing individual RVD (repeat variable diresidue) flanked by *Bsa*I sites and generating different combinations of 5' and 3' 4bp-overhangs upon *Bsa*I digestion. These plasmids contain a tetracycline resistance gene (*TetR*) for selection in E. coli.

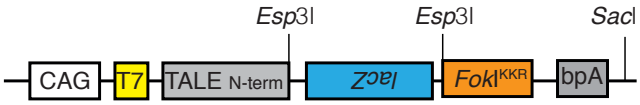
-13 *Bsa*I-based intermediate recipient vectors, designed to accommodate up to 10 RVDs in one Golden Gate reaction (pFusA, pFusA30A; pFusA30B, pFusB1-10). These plasmids contain a Spectinomycin resistance gene (*SpectR*) for selection in E. coli and a *lacZ* cassette for white/blue screening. The RVD array can be released with *Esp*3I generating 4-bp overhangs.

-5 *Esp*3I-based vectors each containing the last half-repeat (pLR-HD; pLR-NG; pLR-NI; pLR-NN; pLR-NK; pLR-NH) necessary to complete each RVD array.

-2 *Esp*3I-based expression vectors containing the *Fok*I nuclease domain. These vectors are designed to accommodate the final RVD array and can be selected in E. coli using Ampicillin.

The destination vectors used in these study are the heterodimeric pCAG-T7-TALEN(Sangamo)-*Fok*I-KKR-Destination and pCAG-T7-TALEN(Sangamo)-*Fok*I-ELD-Destination (**Figure m4**) generated in Dr. Pawel Pelczar's laboratory (UZH, Zurich). CAG promoter drives the expression of the TALE-*Fok*I fusion. These vectors contain a bovine poly-adenilation signal and a T7 promoter upstream the ORF to be used for *in vitro* transcription. A *Sac*I site downstream the TALEN sequence can be used for plasmid linearization.

pCAG-T7-TALEN(Sangamo)-*Fok*I-KKR-Destination (6.4kb) - Addgene #40131



pCAG-T7-TALEN(Sangamo)-*Fok*I-ELD-Destination (6.4kb) - Addgene #40132

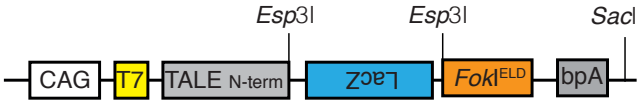


Figure m4: map of the TALEN heterodimeric destination vectors. The CAG promoter can be used for TALEN expression in mammalian cells. The T7 promoter allows *in vitro* transcription of RNA from *Sac*I-linearized plasmid.

The heterodimeric *Fok*I-KKR nuclease includes I538K, E490K, and H537R mutations that prevent the homodimeric activity as well as "Sharkey" mutations S418P and K441E, which enhance the *Fok*I activity. The heterodimeric *Fok*I-ELD nuclease includes Q486E, I499L and N496D mutations as well as "Sharkey" mutations S418P and K441E (Guo et al. 2010). Detailed maps and sequences of all plasmids can be obtained on Addgene: addgene.org/TALeffector/goldengateV2. Detailed information regarding all *Fok*I domain variants is available at: eendb.zfgenetics.org/util-foki.php

3.1.5 CRISPR/Cas9 vectors

hCas9 plasmid (Mali et al. 2013) contains a human codon-optimized Cas9 sequence from *Streptococcus pyogenes* under control of the CMV Enhancer and promoter. A SV40 NLS is fused with the Cas9 C-term (**Figure m5**).

hCas9 (9.4kb) - Addgene #41815

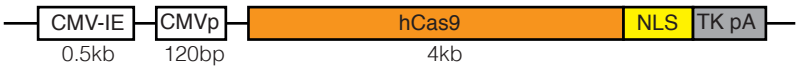


Figure m5: hCas9 vector used in this study, reported in Mali et al. 2013.

Bioinformatic prediction using cNLS Mapper (nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi, Kosugi et al. 2009) indicates that a second, internal monopartite NLS is present at aa 647-660. The sequence of hCas9 vector can be obtained from Addgene: addgene.org/41815.

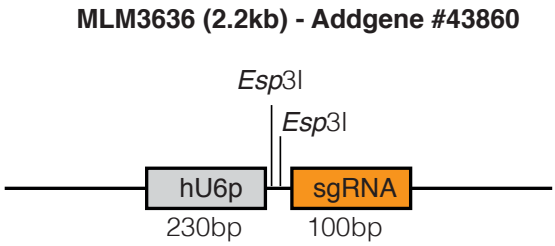


Figure m6: sgRNA expression vector used in this study. The sgRNA vector MLM3636 can be easely modified thanks to the two *Esp*3I sites downstream the hU6 promoter sequence.

MLM3636 (**Figure m6**) contains the invariable sgRNA sequence under the control of hU6 promoter. The sequence of the MLM3636 plasmid can be obtained on Addgene: addgene.org/43860.

3.1.6 Classic restriction-ligation cloning

Restriction-based cloning was performed using standard cloning techniques (Sambrook, Fritsch, & Maniatis, 1989). The amount of 20 μg of plasmid DNA are digested with 20 U of restriction enzyme for three hours at +37ºC in a 50 μl reaction. If necessary, the restriction enzyme is heat-inactivated at +80ºC for 15' and then treated with 10 U of rAPID Phosphatase. Ligation reactions are assembled in 10 μl volume containing 50 ng of vector DNA and insert DNA in a 1:4 molar ratio. Ligations are incubated O/N in a water bath at +16ºC. The Wizard SV Gel and PCR Clean-up System (Promega) has been used for the purification of DNA fragments from agarose gel or from PCR and other enzymatic reactions. Restriction enzymes, T4 DNA ligase and rAPID Phosphatase were purchased from Roche and New England Biolabs and used according manufacturers.

3.1.7 Site directed mutagenesis

Site-directed mutagenesis of the CTCF binding motif in the *Tyr* 3' core element was conducted with the GeneArt® Site-Directed Mutagenesis System (Invitrogen), according to the supplier's indications. Briefly, a set of overlapping mutagenic primers (**Appendix 9.1.2**) with centrally located mutations sites served to PCR-amplify a previously methylated plasmid that contained the sequence to mutate. Recombination between the ends of the PCR product was promoted *in vitro*, and circularized mutated DNA was transformed into DH5α^{TM}-T1^{R} chemically competent bacteria. These bacteria are positive for *McrBC*, an endonuclease that

digests methylated DNA. This property allowed the removal of the methylated plasmid with the original wild-type sequence, leaving only the mutated product, which was purified and checked by restriction digestion and sequence analyses.

3.1.8 T-A cloning

pGEM-T Easy vector (Promega) was used for TA-cloning of PCR products for sequencing purposes, as indicated by the manufacturer. Transformants were spread on LB-agar plates containing 50 μg/ml Ampicillin, 0.1 M IPTG and 40 μg/ml Xgal. Several independent clones are sequenced using universal T7 and SP6 primers.

3.1.9 Golden Gate Cloning

Golden Gate cloning is a technique that allows precise assembly of up to ten DNA sequences in a single reaction (Engler et al, 2009). It is based on Type IIs restriction enzymes, such as *Bsa*I or *Esp*3I, that cut DNA 1-2 bp away from their recognition site and leave a 4 bp-overhang, that can be imposed by the user (**Figure m7**). Thus, unique overhangs can be assigned to the 5' and 3' termini of each DNA fragment. By assigning compatible overhangs to adjacent DNA fragments, complex constructs can be designed and assembled.

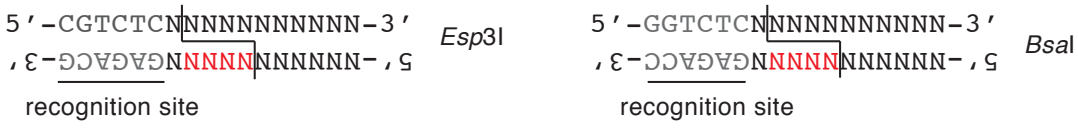


Figure m7: Type IIs restriction enzymes. Recongnition site and distal overhangs generated by type IIs restriction enzymes *Esp*3I and *Bsa*I, used for Golden Gate cloning.

After digestion, each overhang will anneal only with its cognate overhang and the intervening fragments will be joined by the T4 DNA ligase. Such a reaction is really efficient because, unlike standard cloning, digestion and re-ligation of the same fragment to its parental DNA molecule is not possible.

All DNA fragments are mixed in one-tube reaction, together with T4 DNA ligase and a type IIs enzyme. *Bsa*I-based reactions were assembled as follows:

Circular plasmids	100 ng each
10x T4 DNA ligase buffer	2 µl
10x BSA	2 µl
T4 DNA ligase (NEB)	1 µl [400 cohesive ends units]
*Bsa*I (NEB)	1 µl [20 units]
H_20	up to 20 µl

*Esp*3I-based reactions were assembled as follows:

Circular plasmids	100 ng each
10x T4 DNA ligase buffer	2 µl
20mM DTT	1 µl
T4 DNA ligase (NEB)	1 µl [400 cohesive ends units]
*Esp*3I (Thermo Scientific)	1 µl [20 units]
H_20	up to 20 µl

Reactions were cycled as follows:

+37°C 5'; +16°C 10' [10 cycles]
+55°C 10'
+80°C 10'

For assembly reactions including more than five circular plasmids, the reaction was treated for 30' at +37°C with 10 U of Plasmid-Safe Dnase (Epicentre), an ATP-dependent Exonuclease, to digest incomplete products and linear concatamers:

Golden Gate reaction	20 µl
25mM ATP	1 µl
Plasmid-Safe DNAse	1 µl [10 units]

3.1.10 Golden Gate Cloning of TALENs

Two rounds of Golden Gate cloning are necessary to assemble one final 20 RVD-long TALEN plasmid (Figure MM6 A). Firstly, two independent *Bsa*I-reactions are performed (Figure MM9 B and section 3.1.9). RVDs 1 to 10 of the desired array are assembled into the intermediate vector pFusA vector. In parallel, RVDs 11 to 19 are assembled into the intermediate vector pFusB10 vector. Before transformation, the Golden Gate reaction is treated with Plasmid-Safe DNAse (Epicentre) as indicated in 3.1.9. Transformants are plated on LB-Agar plates

containing 100 μg/ml Spectinomycin, 0.1 M IPTG and 40 μg/ml. Three white colonies for each construct are analysed by colony PCR (**Figure m8 A**) with the primers pCR8_F1 and pCR8_R1 (5'-TTGATGCCTGGCAGTTCCCT-3' and 5'-CGAACCGAACAGGCTTATGT-3') and by restriction digest (**Figure m8 B**) with *Afl*II and *Xba*I.

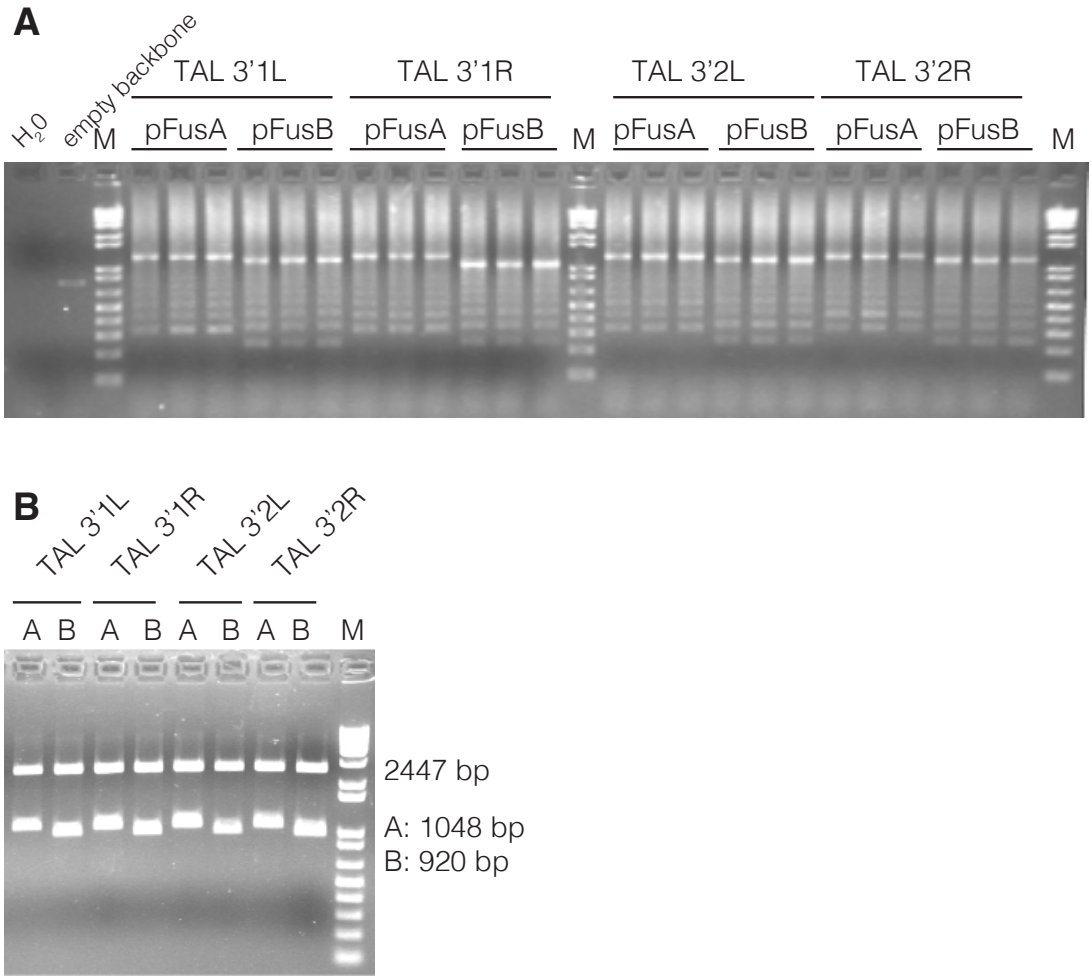


Figure m8: TALEN Golden Gate cloning reaction 1. **A** PCR analysis of the colony resulting from the transformation of the TALEN Golden Gate reaction 1. **B** Positive clones are confirmed by *Afl*III/*Xba*I restriction digest.

To clone the full RVD array into the *Fok*I expression vector, the pFusA and pFusB vectors are assembled into an *Esp*3I-based destination vector, along with the pLR plasmid, encoding the last half-RVD of the array and a linker peptide connecting the RVD array and the *Fok*I domain (**Figure m9 C**).

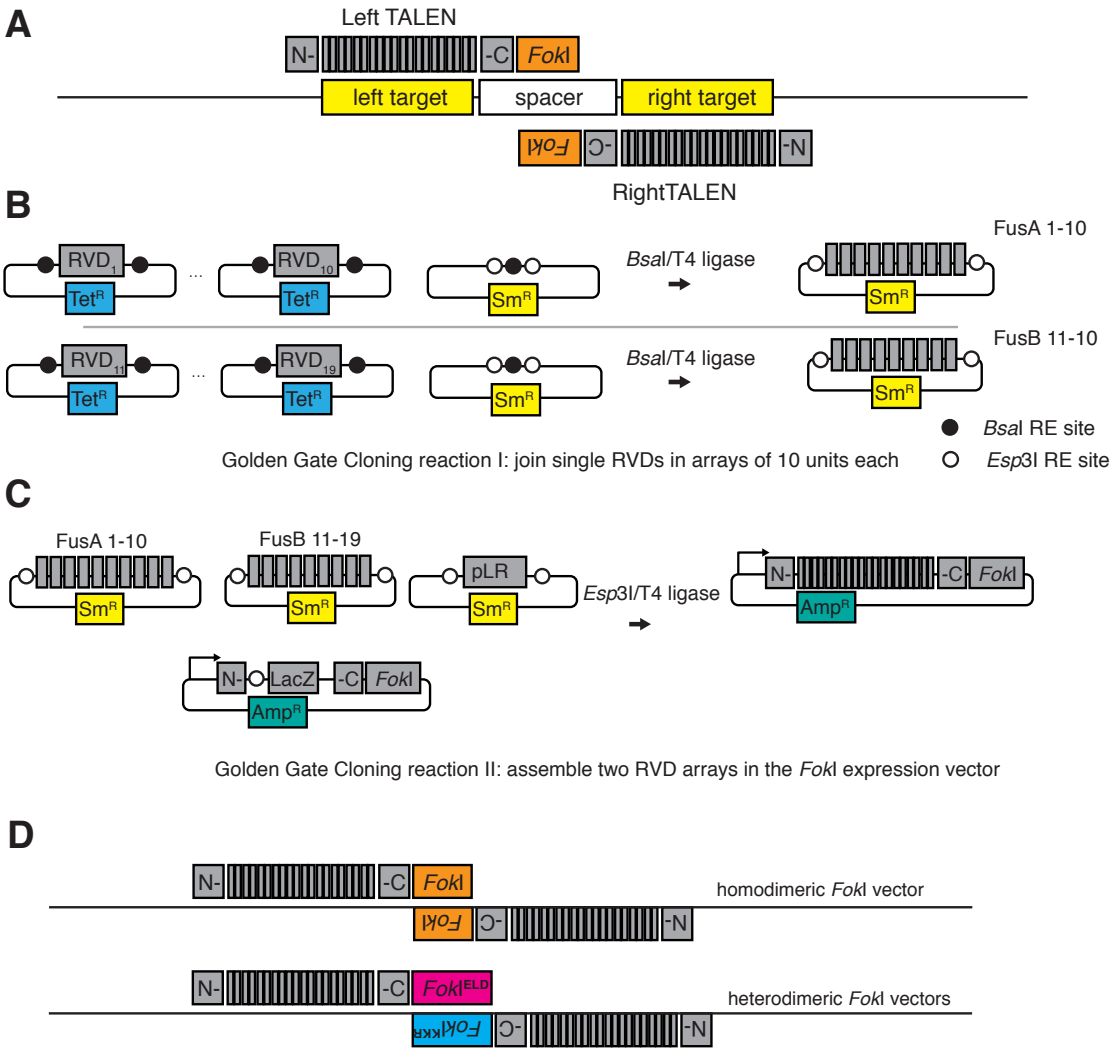


Figure m9: Principle of TALEN Golden Gate Cloning. **A** A TALEN consists in a pair of TAL arrays, each of them fused to a *Fok*I nuclease domain. Each TAL array binds to a target sequence. Two TAL arrays bind distal sequeces separated by a spacer sequence 10-30 bp large. **B** A first reaction is set to assembly intermediate vector containing 10 repeats, each capable of binding one base pare of the intended target. Individual RVDs are assembled in arrays of 10 units by a *Bsa*I-GGC reaction. **C** The intermediate array vectors, assembled in B, are then joined with their halves and transferred to a TALEN destination vector by a *Esp*3I-GGC reaction. **D** Different destination vectors carrying homodimeric or heterodimeric *Fok*I domain can be used.

The second golden gate reaction is performed as indicated in 3.1.9. No Plasmid Safe treatment is necessary in this step. Transformants are plated on LB-Agar plates containing 50 μg/ml Ampicillin, 0.1 M IPTG and 40 μg/ml Xgal. Three colonies for each construct are analysed by colony PCR **(Figure m10 A**) with the primers TAL_F1 and TAL_R1 (5'-TTGGCGTCGGCAAACAGTGG-3' and 5'-GGCGACGAGGTGGTCGTTGG-3'). Two selected clones are analysed by restriction digest with *Bam*HI and *Stu*I (**Figure m10 B**).

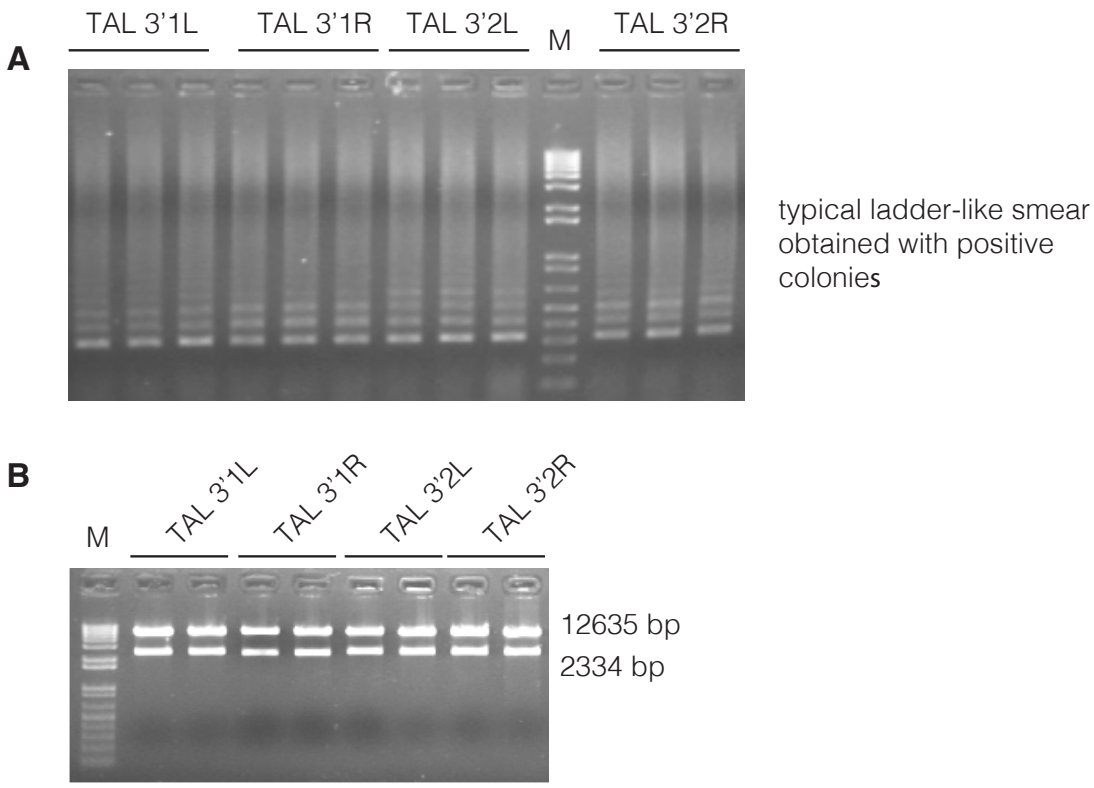


Figure m10: TALEN Golden Gate Cloning reaction 2. **A** Typical result of a colony PCR of the final TALEN golden gate reaction. A smear/ladder is observed in all positive samples. **B** *Bam*HI + *Stu*I restriction digest of clones resulting from the second TALEN golden gate reaction.

Typically, more than 95% of the clones show the expected pattern at both PCR and restriction digest. All constructs are verified by sequencing using the same primers indicated above.

3.1.11 Golden Gate Cloning of CRISPR sgRNA vectors

Two *Esp*3I sites in between hU6 promoter and the invariable sgRNA sequence allow Golden-gate cloning of the sgRNA target sequence in the MLM3636 vector (**Figure m11**). The 20 bp sgRNA target sequence (not including the PAM motif) is

flanked by MLM363-compatible *Esp*3I overhangs and ordered as a single strand oligonucleotide, one for the forward and one for the reverse DNA strand, according to the following design:

-Forward oligonucleotide: ACACC-*target sequence*-G, 26-mer

-Reverse oligonucleotide: AAAAC-*reverse complement of the target sequence*-G, 26-mer

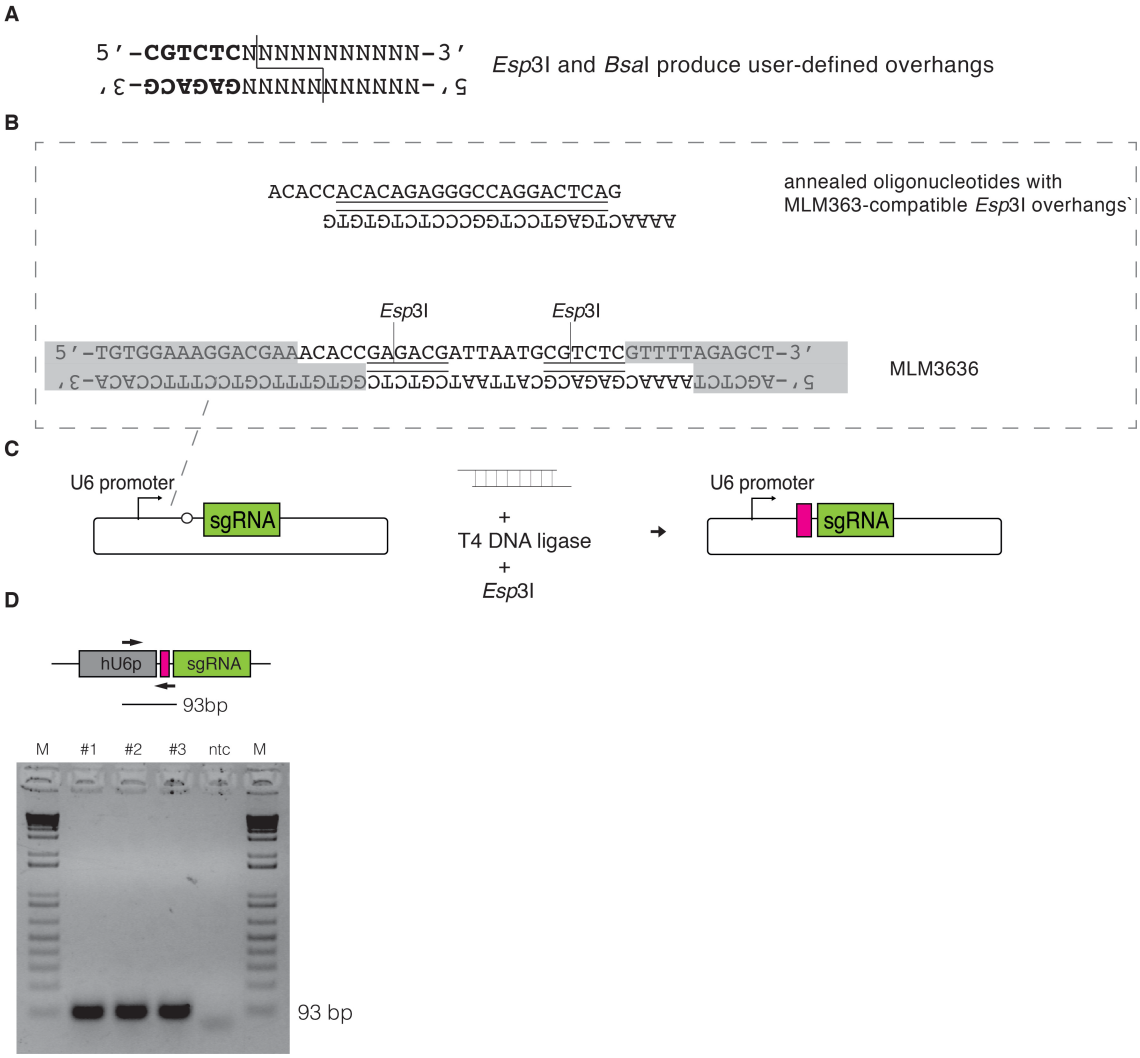


Figure m11: Golden Gate cloning of CRISPR sgRNA vectors. **A** Example of oligonucleotides-based golden gate cloning into the empty MLM3636 vector. **B** Detail of the annealed oligonucleotides carrying overhangs that are compatible to the MLM3636 vector. **C** Diagram of the cloning reaction. **D** Colony PCR of sgRNA vector obtained by Golden-gate cloning. A 93 bp band is expected when the annealed oligonucleotide is correctly inserted in between the *Esp*3I sites of MLM3636 empty vector.

The Forward and Reverse oligonucleotides are mixed in 100 µl H_20 to a final concentration of 10 µM of each oligonucleotide. The oligonucleotide mix is melted

at 95ºC for 5' and annealed 30' at RT. 1 μl of the annealing reaction is Golden Gate cloned in MLM3636 as follows:

Circular MLM3636	100 ng	
Annealed oligonucleotide mix	1 μl	
10x T4 DNA ligase buffer	2 μl	
20 mM DTT	1 μl	
T4 DNA ligase (NEB)	1 μl	[400 cohesive ends units]
*Esp*3I (Thermo Scientific)	1 μl	[20 units]
H_20	up to 20 μl	

Transformants are plated on LB-Agar plates containing 50 μg/ml Ampicillin. Three colonies are analysed by colony-PCR with LKO_1 5' as forward primer (5'-GACTATCATATGCTTACCGT-3') and using the reverse oligonucleotide of the golden gate reaction as reverse primer (**Figure m11 D**). This procedure has been published and reported in detail in Harms et al. 2014.

3.1.12 Bioinformatic tools

Snapgene Viewer (GSL Biotech LLC) was used to annotate plasmid and genomic DNA sequences, to analyse chromatograms and to build plasmid maps. *bl2seq* tool (BLAST/blastn, NCBI) and BLAT (UCSC) were used to align sequences. Primers were designed using PRIMER3 (BioTools, biotools.umassmed.edu/bioapps/primer3) and PRIMERBLAST (NCBI). Molar ratio for restriction-ligation cloning were calculated usign Biomath Calculator (Promega). TALEN Targeter was used to design TAL arrays (tale-nt.cac.cornell.edu/node/add/talen). CRISPR DESIGN (crispr.mit.edu) was used to search for sgRNA binding site and to predict potential off-target sequences. *Bowtie* (Langmead et al. 2009) was also used to select the most unique sgRNA sequences. All mouse genomic sequences were retrieved from the NCBI37/mm9 mouse genome assembly and browsed using UCSC Genome browser (genome-euro.ucsc.edu) and Ensembl (ensembl.org).

3.2 Enhancer-Blocking assay in HEK 293 cells

3.2.1 Plasmid preparation

All constructs used for Enhancer-blocking assay are described in section 3.1.1 and in Lunyak et al. 2007. Prior transfection, plasmid preparations (Genopure plasmid MIDI kit, Roche) are linearized and purified with Wizard SV Gel and PCR-cleanup kit (Promega). Briefly, 20 μg of each plasmid are linearized with *Asp*718I (Roche), with exception of the control plasmids containing the chicken II/III 5'HS4 element, which are linearized with *Apa*I (Roche). A plasmid carrying the *lacZ* coding sequence under the control of the CMV promoter, pZ (Hall et al. 1983), is used for transfection efficiency normalization. pZ is linearized using *Sca*I (Roche).

3.2.2 Cell culture and transfection

Human HEK 293 cells (Shaw et al. 2002) were grown in DMEM medium (Dulbecco's Modified Eagle Medium, Gibco) supplemented with sterile-filtered 10% fetal bovine serum (FBS, Sigma-Aldrich), 2 mM L-glutamine (Invitrogen) and 10 mM HEPES pH 7.4 (Invitrogen) under aseptic conditions using a sterile hood (Telstar Bio II Advance). Cells were cultured at +37ºC, 95% of humidity and 5% CO_2 (CO_2 water-jacketed incubator, Forma Scientific) in plates of different sizes according to the number of cells required for each experiment: 24-well plates (2 cm^2) or p100 (58.10 cm^2) dishes were purchased from Falcon. Usually, cells were grown to 80-90% confluency before they were detached by incubation with 0.25% trypsin-0.02% EDTA (Gibco) for 5 min at +37ºC, and reseeded at 1:8 to 1:10 dilution. Whenever it was necessary to determine the cell number, a cell counting chamber (Sigma) was used according to the manufacturers' specifications. Some 24 hours before transfection, cells are seeded at a density of 2 x 10^5 cells/well (4 x 10^5 cells/ml) in a 24-well plate, using DMEM complete medium. Transfection is performed using Lipofectamine 2000 (Life Technologies) diluted in OptiMEM (Life Technologies) as indicated by the manufacturer. 660 ng of each construct is transfected along with 140 ng of the β-galactidase vector pZ. Each transfection is performed in three technical replicates. At least two biological replicate per experiment are performed.

3.2.3 Preparation of cellular extracts

About 24 hours post-transfection, cells are gently washed with ice-cold PBS and lysed in 125 μl 1x Reporter Lysis Buffer (Promega), with the help of a rubber policeman. Lysates are vortexed and spun 2 min at 12'000 x g in a refrigerated centrifuge. Cleared lysates are transferred to a fresh tube and kept on ice. A 1:100 dilution is prepared in lysis buffer. The remaining volume is stored at -80°C.

3.2.4 β-galactosidase activity measurement

To measure β-galactosidase activity we used the the hydrolysis of the chromogenic substrate o-nitrophenyl-b-D-galactoside (ONPG, Sigma Aldrich), as reported by Hall (Hall et al. 1983). First, 100 μl of a 1:100 dilution of each lysate are mixed with 400 μl of Z-buffer (100 mM Na_2HPO_4 pH 7.2, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol) and 100 μl of ONPG (4 mg/ml in 100 mM Na_2HPO_4 pH 7.2). Plain lysis buffer was used as negative control to blank the microplate reader. Samples are incubated at +37°C for 30' and observed at 10' intervals. As soon as a pale yellow colour is observed, the reaction is stopped with 250 μl 1M Na_2CO_3. Then, 200 μl of the reaction are transferred to a transparent 96 well-plate (Nunc) and 414 nm absorbance is measured in a microplate reader.

3.2.5 Luciferase activity measurement

A volume of 100 μl of a 1:100 lysate dilution of each sample is transferred to a opaque-white 96-well plate (Berthold). Luminescence is measured in an Orion Microplate Luminometer (Berthold), programmed to inject 50 μl Luciferin (Promega) in each well, with an exposure time of 10 seconds and a well-to-well delay of 2.05 seconds.

3.2.6 Data analysis

Luciferase activity was corrected for transfection efficiency using the β-galactosidase activity, obtaining a value of Relative Luciferase activity. This value is then normalized for the amount of transfected DNA (pmol). Enhacer-blocking activity is obtained dividing the Relative Luciferase activity of pELuc by that of each sample. This ratio is shown along with the standard deviations of the triplicates.

3.3 Enhancer-blocking Assay in zebrafish embryos

Experiments with zebrafish embryos were carried out in Dr. José Luis Gómez-Skarmeta's laboratory, at the CABD, Seville, with the help of Dr. Ana Fernández-Miñán. All procedures met the EU animal research guidelines (Directive 2010/63/UE) and the Spanish animal research legislation (L32/2007; RD53/2013; L6/2013) and were approved by the CSIC and local Ethics Committee.

3.3.1 Plasmid purification

All constructs used for EBA in zebrafish are described in section 3.1.2. Plasmid DNA preparations are obtained using Genopure plasmid MIDI kit (Roche). To remove traces of RNAse, 20 μg of DNA are diluted in water in a total volume of 400 μl. 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) is added. The sample is vigorously mixed and spun 5' at 20'000 x g at RT. The superior, aqueous phase is transferred to a fresh tube. An equal volume of chloroform:isoamyl alcohol is added and the aqueous phase is collected after the same procedure. Two volumes of ice-cold absolute ethanol are used to precipitate the DNA. The sample is centrifuged 15' at 20'000 x g at +4°C. The DNA pellet is then washed with 400 μl of 70% ethanol, air-dried and resuspended in RNAse-free water.

3.3.2 Microinjection into zebrafish embryos

Transient transgenic zebrafish embryos were generated by *Tol2*-based transgenesis (Kawakami et al. 2009). p48RCAR-based construcs (3.1.2) were

microinjected into one-cell stage zebrafish embryos (AB strain) using a IM-300 microinjector (NARISHIGE) as previously reported (Bessa et al. 2009). Plasmid DNA was injected at a concentration of 40 ng/ul along with 50 ng/ul of *Tol2*-transposase mRNA. Phenol red (0.05%) was added to the microinjection mix in order to track visually the success of the injection. The microinjection mix was delivered at the yolk/cell interphase. After microinjection, embryos were kept in E3 medium (5 mM NaCl, 0.17 mM KCl, 0.4 mM $CaCl_2$, 0.16 mM $MgSO_4$) at +28°C. No more than 50 individuals were kept in each plate. After 24 hours, all plates are inspected for the presence of dead or deformed embryos, which are removed. The medium is replaced with fresh E3 medium supplemented with 0.003% PTU (1-Phenyl-2thiourea, Sigma) to inhibit the development of pigmentation that would quench the GFP fluorescence.

3.3.3 Imaging

Embryos were imaged at 36 hfp (hours post-fertilization) using an Olympus SZX16 fluorescence microscope and the CellSens Entry 1.6 software (Olympus). Briefly, zebrafish embryos were dechorionated using two pairs of fine-tip tweezers and anesthetized by adding several drops of 0.4% tricaine (Sigma-Aldrich) to the E3 medium. All embryos were imaged using the same settings.

3.3.4 Image processing and data analysis

LasexPix image analysis software (Bio-Rad) was used to measure the GFP+ areas in zebrafish embryos fluorescence images. Fluorescent areas observed in midbrain or somites are manually divided and summed for each individual. A Median, non-parametric test is used to assess the statistical significance of the differences observed between experimental groups.

3.4 Chromosome Conformation Capture (3C)

Chromosome conformation capture technology allows the detection of physical, long-range interaction between distal DNA sequences that can be used to infer the three-dimensional chromatin organization of a given locus. The current procedure is based on Hagege et al. 2007 and was set up with the help of Dr.

Ana Fernández-Miñán and Dr. Jose Luis Gomez-Skarmeta, CABD, Seville, Spain.

3.4.1 Cell culture

Mouse melanoma cell line B16-F1 and mouse fibroblast NIH-3T3 L929 cell line were grown in DMEM medium (Dulbecco's Modified Eagle Medium, Gibco) supplemented with sterile-filtered 10% fetal bovine serum (FBS, Sigma-Aldrich), 2 mM L-glutamine (Invitrogen) and 10 mM HEPES pH 7.4 (Invitrogen) under aseptic conditions using a sterile hood (Telstar Bio II Advance). Cells were cultured at +37ºC, 95% of humidity and 5% CO_2 (CO_2 water-jacketed incubator, Forma Scientific) in T175 flasks (Falcon). Usually, cells were grown to nearly 100% confluence before they were detached by incubation with 0.25% trypsin-0.02% EDTA (Gibco) for 5 minutes at 37ºC, and reseeded at 1:8 to 1:10 dilution. Whenever it was necessary to determine the cell number, a heamocytometer (Sigma) was used according to the manufacturers' specifications.

3.4.2 RNA extraction and quantification of *Tyr* expression

RNAeasy kit (QIAGEN) was used to extract RNA from 1 x 10^6 cells, as indicated by the manufacturer. Two μg of RNA were treated with RNAse-free DNAse (Roche). Subsequently, cDNA was obtained from 1 μg of RNA using Superscript III (Life technologies) and randon hexamer oligonucleotides (Life Technologies) as indicated by the vendor. One μg of RNA was kept without reverse transcription as a control for DNA contamination. A 1:20 dilution of cDNA was used for RT-qPCR using gene-specific TaqMan Gene Expression Assays (Applied Biosystems) in a ABI 7500 machine. qPCRs were conducted in triplicates and repeated twice. Mm00495817_m1 probe was used for mouse *Tyr* expression. Mm00446973_m1 probe was used for mouse TBP (TATA binding protein) expression, for normalization (Gimenez et al. 2003; Lavado et al. 2006).

3.4.3 3C Primer design

Primers were designed using Primer3 and PrimerBLAST, 70-150 bp away from a *Dpn*II site flanking fragments of interest (promoters, CTCF-binding sites…) and ''neutral'' sequences, included as control. Each oligonucleotide is design outward

looking, pointing to the closer restriction site (**Figure m12 A**). The specificity of a primer is evaluated using PrimerBLAST and Ensembl BLASTN optimized for near-exact matches. Nevertheless, digestion and re-ligation process produce a scrambled, shuffled genome that current algorithms are not capable of predicting. PCR validation is therefore the golden standard.

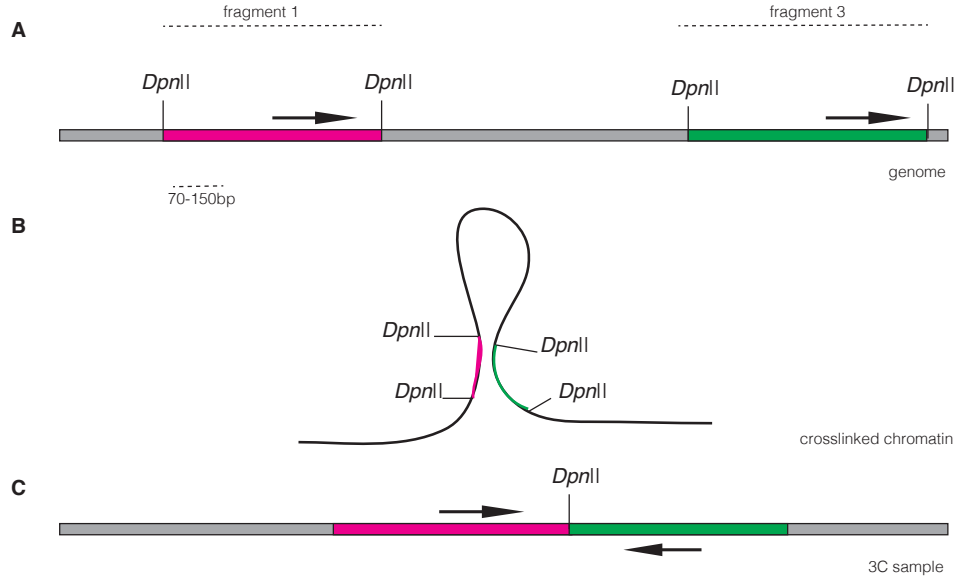


Figure m12 3C primer design. **A** Primers are design 70-150 bp away from the chosen restriction site, outward-looking. In the given example, primers are designed not yielding any amplification in the untreated genome. **B** In the 3C procedure, sequences that are in physical proximity are cross-linked. **C** Only upon digestion and religation, fragment 1 and 3 eventually came in contact and the designed oligonucleotide can anneal and yield a PCR product.

3.4.4 Production of BAC control template

A template containing all 3C chimeric ligation products is produced for absolute-qPCR calibration (**Figure m13**). Two BACs covering the region object of the study (3.1.3) where obtained from CHORI (bacpac.chori.org/). Equimolar amounts of each BAC are mixed to a total of 2 μg and digested o/n using 50 U of *Dpn*I (Roche) in 500 μl of 1x *Dpn*I buffer. After digestion, a 20 μl aliquot is kept for electrophoretic analysis. The rest of the volume is phenol-extracted, ethanol-precipitated and resuspended in 88 μl of mqH_20. Ligation is performed with 2 μl (6 Weiss U) of T4 DNA ligase (Promega) in 100 μl volume in a thermocycler, using the following program: +4°C, 3 hours; +14°C, 3 hours (4x). A 4 μl aliquot is kept for electrophoretic analysis. Next, 80 μl H_20 and 20 μl NaAc 3M pH 5.2 are added

to the rest of the reaction. The mixture is phenol-extracted, ethanol precipitated and resuspended in 100 µl H_20.

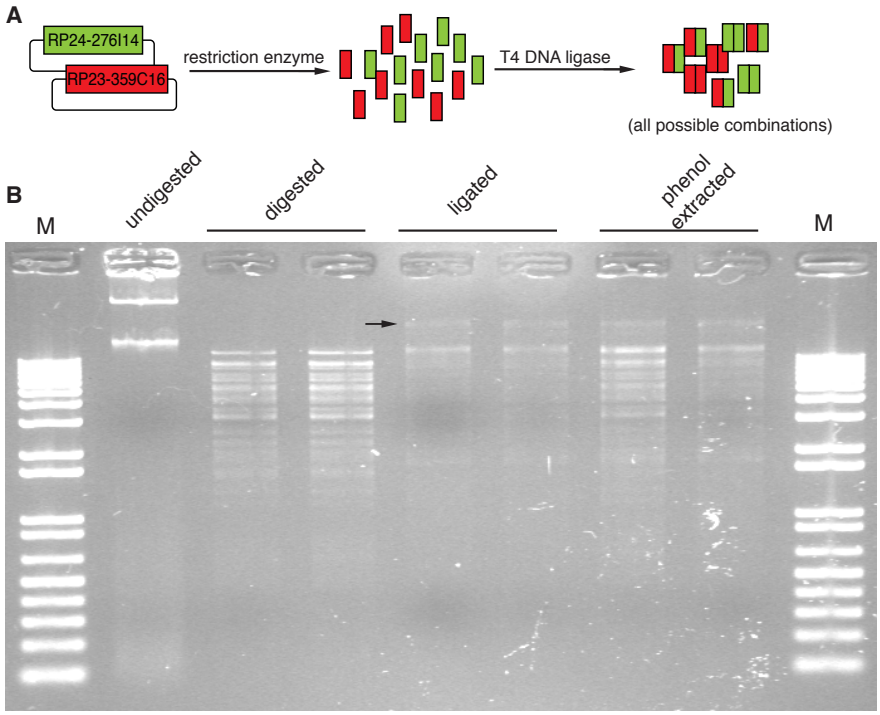


Figure m13 Production of 3C BAC control template. **A** Two BACs are mixed in equimolar amount, digested and religated in order to produce a library of all possible chimeric ligation products. **B** After ligation, a high-molecular weight band is observed, indicating the success of the digestion-ligation reaction.

3.4.5 Cross-link and isolation of intact nuclei

Confluent cells (80-90%) are washed with PBS, collected by trypsinization and counted using a haemocytometer (Sigma). Ten million cells are resuspended in 5 ml of PBS in a 15 ml falcon. 5 ml of filtered, 4% PFA/PBS (Merk) are added and the solution is incubated 10 min at room temperature with continuous agitation in order to prevent the formation of cell clumps. Fixation is quenched by adding 1.425 ml of ice-cold 1M glycine (Calbiochem). The sample is spun 8 min at 1300 rpm in a refrigerated centrifuge. The pellet is then resuspended in 5 ml ice-cold lysis buffer (10 mM Tris HCl, 10 mM NaCl, 0.3% NP40 supplemented with fresh protease inhibitors, 1x Complete, Roche) and incubated on ice 10 min with gentle agitation. The sample is then centrifuged 5 min at 1300 rpm in a refrigerated centrifuge, resuspended in 500 µl lysis buffer and transferred to a 1.5 ml tube. With the help of a plastic pestle the cells are homogenized on ice every 10 minutes until complete lysis. MGP staining (methyl green pyronin, Sigma Aldrich)

is used to follow the lysis of cellular membranes and the release of intact nuclei. Complete lysis is obtained when the cytoplasm, stained in pink, in no longer visible by MGP staining.

3.4.6 Chromatin digestion and ligation

Intact nuclei are spun 5' at 2600 rpm at 4ºC and resuspended in 485 µl of 1x restriction digestion buffer (NEB). This step is repeated to remove traces of lysis buffer that may inhibit the restriction enzyme. Fifteen µl of 10% SDS are added and the sample is incubated 1h at +37ºC with shaking. To precipitate the SDS, 55 µl of 20% Triton X-100 are added. The sample is incubated 1h at 37ºC with shaking. A 10 µl aliquot is kept as undigested control. Then, 300 U of restriction enzyme (NEB, Roche) are added and the sample is digested o/n at +37ºC with shaking. After digestion, a 10 µl aliquot is kept as digested control. Digestion efficiency was determined at this point. The control aliquots were taken up to a volume of 95 µl, using 10 mM TrisMHCl pH 7.5. Next, 5 µl of 10 mg/ml proteinase K (Roche) were added to each sample. They were incubated at 65ºC for 1 hour to reverse the cross-links and the DNA was purified by phenol:chloroform:isoamyl alcohol extraction. The purified undigested and digested controls for each cell type were run on a 0.6% agarose gel and compared. If digestion had taken place, the samples were ready for the next step.

The restriction enzyme is heat-inactivated by incubating 25 min at +65ºC. The digested chromatin is then transferred to a 15 ml falcon and diluted in 7 ml of 1x T4 DNA ligase buffer (30 mM TrisHCl, 10 mM $MgCl_2$, 10 mM DTT, 10 mM ATP). Fifteen µl of T4 DNA ligase (Promega, 15-45 Weiss Units) are added. The sample is incubated o/n at +14ºC. To determine ligation efficiency, aliquots of 100 µl were taken and processed as in the previous step. Purified ligated samples were run along undigested and digested controls and compared. If ligation had occurred as expected, the samples were further processed.

3.4.7 Cross-link reversal and DNA purification

Digested, re-ligated chromatin is incubated o/n at +65ºC, after the addition of 30 µl of 10mg/ml Proteinase K (Roche), to reverse the DNA cross-link and facilitate DNA extraction. Next, 30 µl 10mg/ml RNAse A (Roche) are added to digest

cellular RNAs. Seven ml of Phenol-Chloroform-Isoamyl alcohol (25:24:1) are added. The sample is vigorously mixed and spun 10 min at 4000 rpm. The supernatant is carefully transferred to a fresh 15 ml tube. An equal volume of Chloroform-isoamyl alcohol (24:1) is added, the sample is then mixed and centrifugated 10 min at 4000 rpm. The supernatant is carefully transferred to a 50 ml falcon. Seven ml mq H_20, 933 μl 3M NaAc pH 5.2 and 24 ml 100% Ethanol are added and the sample is centrifuged 45 min at 4000 rpm at +4ºC. The resulting pellet is washed in 70% ethanol, air-dried and resuspended in 150 μl TE (10 mM TrisHCl, 1 mM EDTA).

3.4.8 Quantification of distal DNA interactions by qPCR

A standard, end-point PCR to detect the conserved interaction at the Ercc3 locus (see Appendix for oligonucleotide sequences) is performed prior to qPCR quantification using three different dilutions of the 3C sample. The success of this PCR is a first indication that the 3C procedure was performed correctly and it is ready for quantification using qPCR. The linear range of amplification was determined by performing qPCR on serially diluted BAC control and 3C sample templates, with different primer pairs. Once the working DNA concentration range had been established, qPCRs were conducted, in triplicates. Enrichment in a given ligation product was taken as a measure of the interaction frequency between two genomic loci. Hence, relative interaction frequencies were calculated by using a standard curve generated from a serial dilution of the BAC mix control template (3.4.4), which contained all possible ligation products in equimolar amounts. To enable the comparison of the results from different 3C samples, the interactions observed in each cell line were normalized to that of the Ercc3 region, an unrelated locus that is often used as control, since it is considered to adopt the same spatial conformation regardless the cell type (De Laat & Grosveld, 2003; Palstra et al., 2003; Drissen et al., 2004). qPCR quantification was performed using SYBR Green PCR Master Mix (Applied Biosystems) using an ABI 7500 Real-Time PCR Machine. Two biological replicates of each cell type were analysed at least twice.

3.5 Transcription-Activators Like Effector Nucleases

3.5.1 TALEN design

Repeat Masker (Tempel, 2012) was used to identify repeat-free sequences in the target region. TALEN Targeter (Doyle et al. 2012, tale-nt.cac.cornell.edu/node/add/talen) was used to identify TALEN binding sites in the selected repeat-free sequences and to obtain the corresponding RVD arrays. We selected DNA spacer length range of 12-23 bp and TAL arrays of 15-20 RVDs each. The RVD used to recognise G was NN.

TALENs with the lowest predicted off-targets were considered and RVD arrays with the highest NG and HD content were selected. The selected TALENs were assembled by Golden Gate Assembly (Engler et al. 2009), using the Voytas Golden Gate TALEN kit (Addgene #1000000024). Moreover, to reduce off-targeting possibility, an obligate heterodimeric FokI mutant was used (Doyon et al. 2011, Addgene #40131 and #40132). Details of the cloning process and of the destination vectors used are available in section 3.1.9 and 3.1.10 of this PhD thesis.

3.5.2 TALEN validation in mouse Neuro2A cells

Combination of TALEN pairs targeting adjacent DNA sequences, that would induce deletion of the intervening DNA, were transfected in mouse Neuro2a cells. Briefly, 1 x 10^6 cells were plated in 6-well plates in 2 ml of complete DMEM medium in the absence of antibiotics. Then, 300 ng of each TALEN plus 100 ng of pEGFP-N1 (Clontech) - to follow the transfection efficiency - were transfected with 6 μl of Lipofectamine 2000 (Life Technologies) as indicated by the manufacturer. Some 72 hours post-transfection, the cells are lysed in 800 μl of a SDS-free lysis buffer (50 mM KCl, 10 mM TrisHCl pH 8, 0.45% NP40, 0.45% Tween-20, 500 ng/mL Proteinase K) to bring DNA in solution. Lysates are incubated for 3 hours at +56°C. The Proteinase K is then heat-inactivated at +95°C for 10 minutes. One microliter of the sample is directly used for PCR, or stored at 4°C.

3.5.3 *In vitro* cleavage assay for TALEN validation

We used an *in vitro* cleavage assay (modified from Jinek et al., Elife 2013) to assess the cleavage activity of purified TALEN proteins on a PCR product containing the corresponding target sequence. Briefly, HEK293 cells were plated at a density of 4x 10^5 cells/ml in a p24 plate, in complete DMEM without antibiotics. The next day, 400 ng of each TALEN (200 ng of each TALEN monomer) plus 50 ng of pEGFP-N1 plasmid were transfected using Lipofectamine 2000 (Life Technologies). Some 72 hours post-transfection, cells were washed with ice-cold PBS and lysed in 250 μl of 1x cleavage buffer (20 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 5% Glicerol, 0.1% Triton X-100, and 1x Roche Complete Protease Inhinitor cocktail). Lysates were collected in a 1.5 ml tube, vortexed 10'' and spun 2' at 12'000 x g at +4°C. Cleared lysates were transferred to a fresh tube.

Ten μl of each TALEN lysate were mixed with 500 ng of a PCR product containing the relative target site, in 1x Cleavage buffer, in a total volume of 20 μl. Samples were incubated at +30°C for 3 hours. Prior to gel electrophoresis, samples were treated with 500 ng Proteinase K for 10' at +65°C. As control, each target PCR product was incubated with a TALEN lysate specific for a different target of with a lysate of mock-transfected cells.

3.5.4 TALENs RNA *in vitro* transcription

TALEN mRNAs (5'-capped and 3'-polyA tailed) were transcribed *in vitro* from plasmid templates, using the mMessage mMachine T7 ULTRA kit (Ambion) as indicated by the manufacturer. Briefly, 20 μg of each TALEN expression vectors (described extensively in section 3.1.4 and 3.1.10) were linearized with 20 units of *Sac*I (Roche) for three hours in 50 μl volume. DNA was purified with Wizard SV Gel and PCR Clean-up system (Promega) and quantified using a Nanodrop device. The successful linearization was also confirmed by agarose gel electrophoresis. One μg of linear DNA was used for T7 RNA polymerase-mediated RNA transcription. The 5'-cap was added in this first step. After DNAse treatment, a poly A tail was added to the RNA with an enzymatic reaction as indicated by the manufacturer. Capped, polyA tailed mRNA was then purified by

gel filtration using Nucaway Spin Columns (Ambion) rehydrated in RNAse-free microinjection buffer (1 mM TrisHCl pH 7.5, 0.1 mM EDTA pH 7.5) (**Figure m14**).

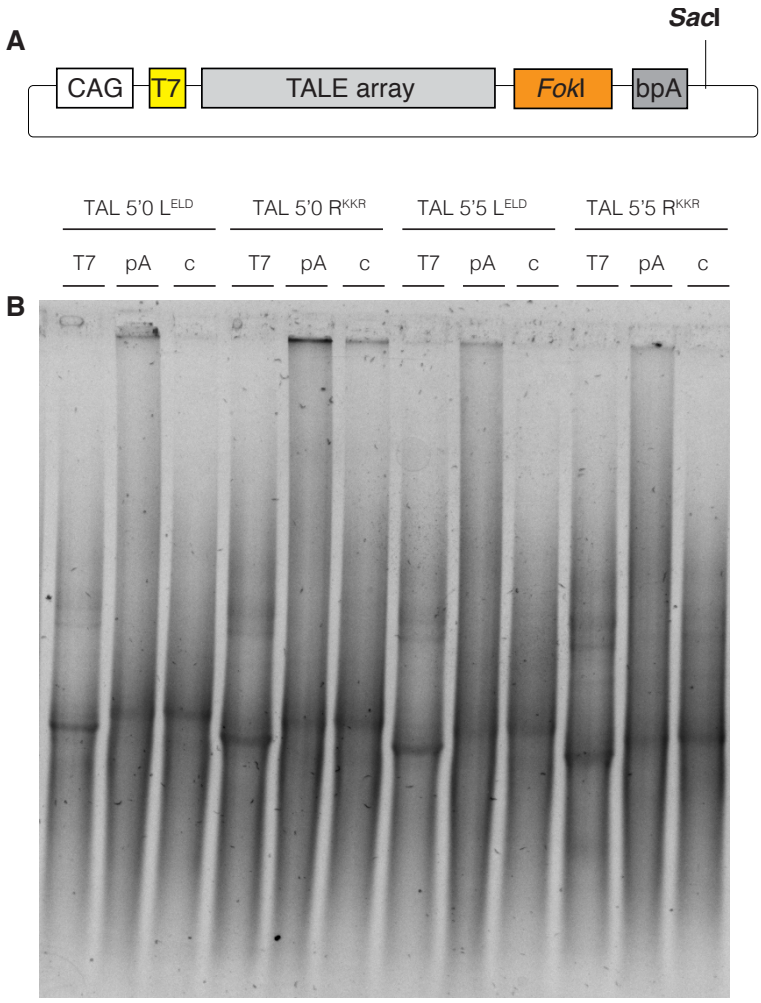


Figure m14: in vitro transcription of TALENs. **A** The *Sac*I-linearized vector, containing the T7 promoter sequence upstream to the TALEN open reading frame, was used as template for *in vitro* transcription. **B** A sample of each step of the transcritpion and tailing is run in a TBE-agarose gel. A clear supershift is observed in the RNA after A-tailing (pA).

Different 2.5 µl aliquots of each step of the RNA production were kept, heat-denatured and run in a TBE-agarose gel. Bands are visualized under UV light after staining with SYBR Safe (Invitrogen). RNA concentration was measured using a Nanodrop device. RNA was next diluted at 200 ng/µl and stored at -80°C in 20 µl aliquots until use. Before proceeding with microinjection into mouse fertilized eggs, TALEN mRNA are mixed and diluted to the proper working concentration in microinjection buffer (1 mM TrisHCl, 0.1 mM EDTA) and centrifuged at 30'000 x g for 1 hour at +4°C, to clear the solution from any particle or debris that may clog the microinjection needle and/or interfere with mouse embryo development.

3.6 CRISPR

3.6.1 CRISPRs design

CRISPR/Cas9 nucleases can virtually target any 20 bp-long DNA sequence that is followed by the PAM (Protospacer Adjacent Motif) motif 5'-NGG-3'. An example of possible CRISPR target sites within the mouse *Tyr* exon 1 is given in **Figure m15**.

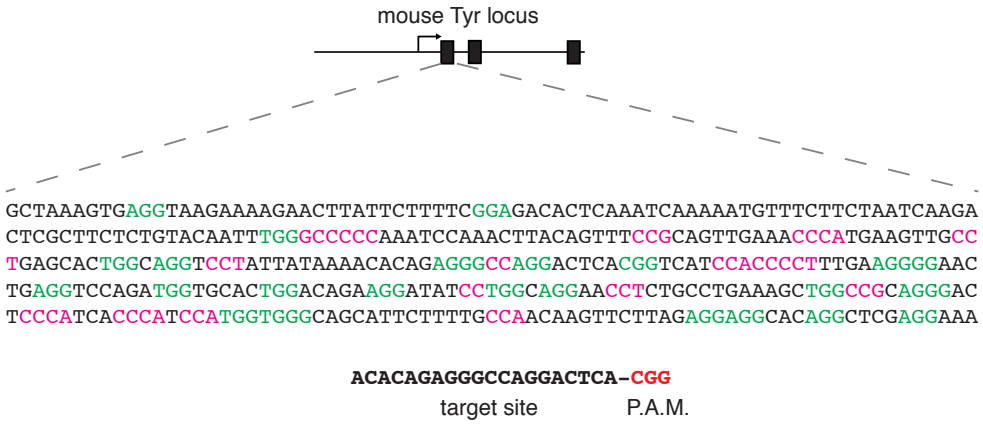


Figure m15: PAM motifs found in the mouse *Tyr* exon. PAMs on the forward strand are indicated in green. PAMs on the reverse strand are shown in purple. One selected sequence is shown below.

For the CRISPRs used in this study, N_{20}-NGG sequences were picked manually or with the help of Zhang Lab CRISPR Design online tool (crispr.mit.edu) in the same DNA region used to design TALENs, previously identified using RepeatMasker (3.5.1). These sequences are then screened to find those that are unique within the mouse genome. CRISPRs have a reported tendency of off-targeting even when mismatches are present between the desired target site and the undesired (Fu et al. 2013; Mali et al. 2013; Pattanayak et al. 2013 and Hsu et al. 2013). Therefore, a ''relaxed'' BLAST analysis is used to detect sequences that display multiple mismatches with the desired target. To this end, a script was compiled using Bowtie and run for each of the possible CRISPR target sequences. An example for the 3'3 CRISPR is given in **Table m1**.

Table m1 *Bowtie* script used to identify potential off-target sites.

```
 ./bowtie -c m_musculus_ncbi37 AGCTCAGTAGAGTACTAGGTAGG -v 3 -l 23 -k 100
0
0         +         135522000 AGCTCAGTAGAGTACTAGGTAGG       IIIIIIIIIIIIIIIIIIIIIII       0         7:G>T,13:C>A
0         -         91016905  CCTACCTAGTACTCTACTGAGCT       IIIIIIIIIIIIIIIIIIIIIII       0         12:C>A,16:C>T,19:G>A
0         +         72174498  AGCTCAGTAGAGTACTAGGTAGG       IIIIIIIIIIIIIIIIIIIIIII       0         12:A>T,15:A>T,16:G>A
```

In this case, the three off-targets found by Bowtie map in intronic regions of the *Stx1a, Nup210* and *Gabrb1*. The potential off-targets show 2, 3 and 3 bases of mismatch, respectively.

Recently, online tools for CRISPR design and off-target prediction became available. The selected CRISPRs were therefore re-analysed using newly available tools design by members of the Zhang laboratory at the MIT, Boston, USA. Those tools can be browsed at crispr.mit.edu. This second analysis confirmed the good quality of the selected targets. Detailed results of the potential off-target search are listed in the result section. CRISPR sequences are ordered as oligonucleotides (Sigma Aldrich), annealed and cloned as described in the section 3.1.9 and 3.1.11 of this PhD thesis.

3.6.2 CRISPR/Cas9 validation in mouse Neuro2a cells

For in vitro tests, mouse Neuro2a cells were plated in a 24-well plate at a density of 4 x 10^5 cells/ml in a 0.5 ml volume of complete DMEM without antibiotics. On the next day, 400 ng of hCas9 plasmid and 100 ng of the sgRNA plasmid were transfected along with 50 ng of pEGFP-C1 as transfection efficiency reporter. Transfection was performed using Lipofectamine 2000 as reported previously. Thereafter, 72 hours post-transfection, genomic DNA was obtained with a SDS-free buffer(50 mM KCl, 10 mM TrisHCl pH 8, 0.45% NP40, 0.45% Tween-20, 500 ng/ml Proteinase K) and used directly for PCR. CRISPRs activity was evaluated by T7 Endonuclease I assay or by PCR using primers designed to detect chromosomal deletions.

3.6.3 T7 Endonuclease I assay

In order to detect nuclease-induced mutations, insertions and deletions (*indels*), we used a mismatch-detection assay based on T7 Endonuclease I digestion of PCR products (**Figure m16**). Briefly, the region surrounding the nuclease target

site is PCR-amplified with primers yielding a PCR product not smaller than 400 bp and not larger than 1 kb.

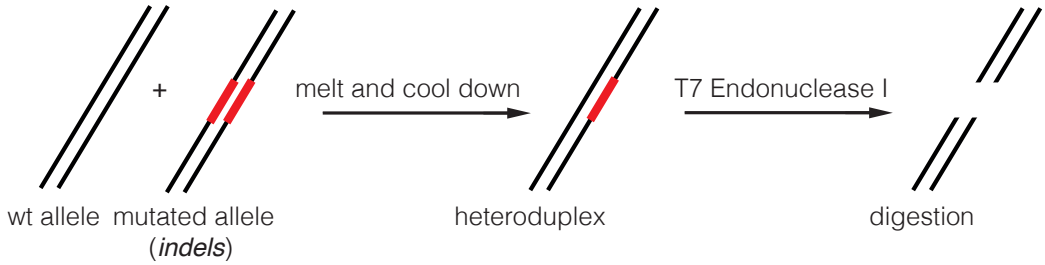


Figure m14: Principle of the T7 Endonuclease I. PCR products are melted and re-annealed to favour the formation of heteroduplexes. The T7 Endonuclease I is used to detect eventual mispairing of wild-type with mutated DNA strands.

To produce heteroduplexes, the PCR product is denatured at +95°C and allowed to re-nature slowly, in a thermocycler, using the following cycling parameters:

+95°C 10'
+95°C -> +85°C -2C/sec
+85°C 1'
+85°C -> +75°C -0.3C/sec
+75°C 1'
+75°C -> +65°C -0.3C/sec
+65°C 1'
+65°C -> +55°C -0.3C/sec
+55°C 1'
+55°C -> +45°C -0.3C/sec
+45°C 1'
+45°C -> +35°C -0.3C/sec
+35°C 1'
+43°C -> +25°C -0.3C/sec

The reaction is prepared as follows:

PCR product	10 µl [out of a 25 µl PCR reaction]
10x NEB buffer 2	2 µl
T7 Endonuclease I, NEB	0.5 µl [5 units]
H_20	7.5 µl

and incubated at +37°C for 25 min. All reaction volume is resolved in a 2% TBE-agarose gel. A typical result is shown in **Figure m17** using a sgRNA designed to target the EGFP coding sequence (Auer et al. 2014).

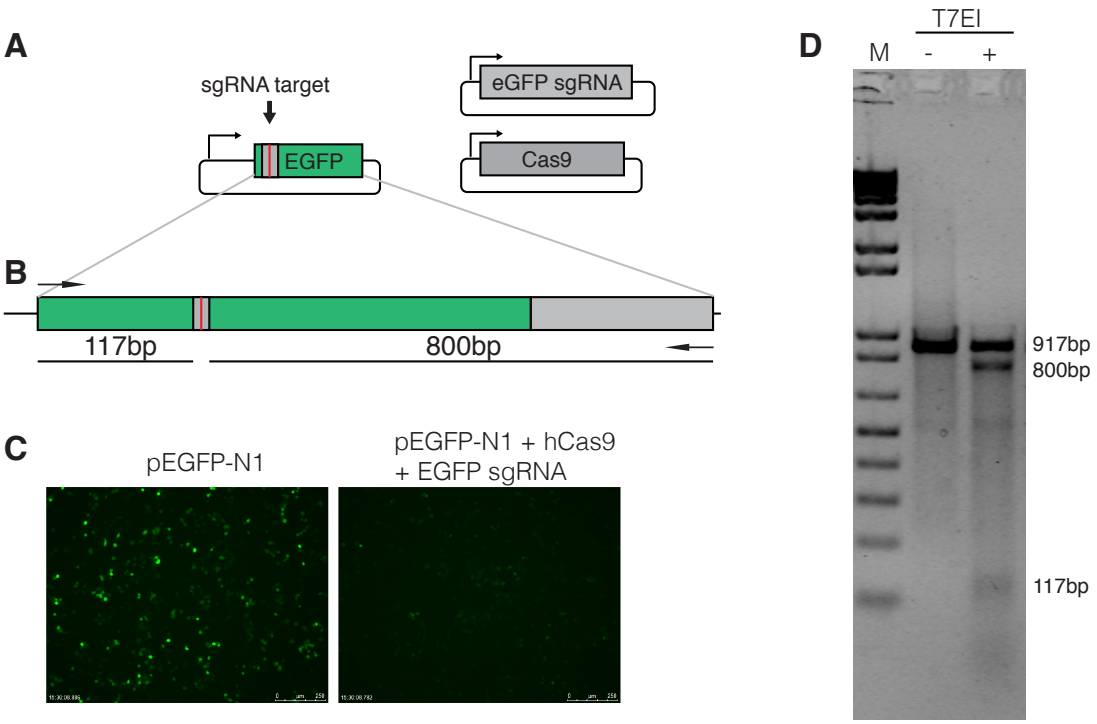


Figure m17 Example of T7 Endonuclease I assay. **A** Mouse Neuro2A cells are transfected with the GFP expression vector pEGFP-N1 (Clontech), along with the hCas9 expression vector and a sgRNA expression vector targeting a sequence within the EGFP open ready frame. **B** EGFP coding sequence is PCR amplified using the primers indicated. The expected nuclease cut site is found 117 bp downstream the forward primer. **C Fluorescence microscopy images of cells transfected with either p**EGFP alone or in combination with CRISPR plasmids. EGFP signal is reduced by CRISPR. **D** T7 Endonuclease I digestion of total DNA extracted from transfected cells. The PCR product is cleaved at the expected position indicating that *indels* have been produced at the CRISPR target site

3.6.4 Cas9 and sgRNA *in vitro* Transcription

Cas9 ORF including NLS was PCR amplified from the vector hCas9 with Expand Long Template PCR system (Roche) with the following conditions:

1. +94°C		2’
{2. +94°C		20”
3. +50°C		30”
4. +72°C		6’ } 30 cycles
5. 72°C	8’	

The T7 RNA promoter sequence was added 5’ to the forward oligonucleotide. One microgram of the purified PCR was used as template for in vitro RNA transcription using the mMessage mMachine T7 ULTRA kit as indicated by the manufacturer.

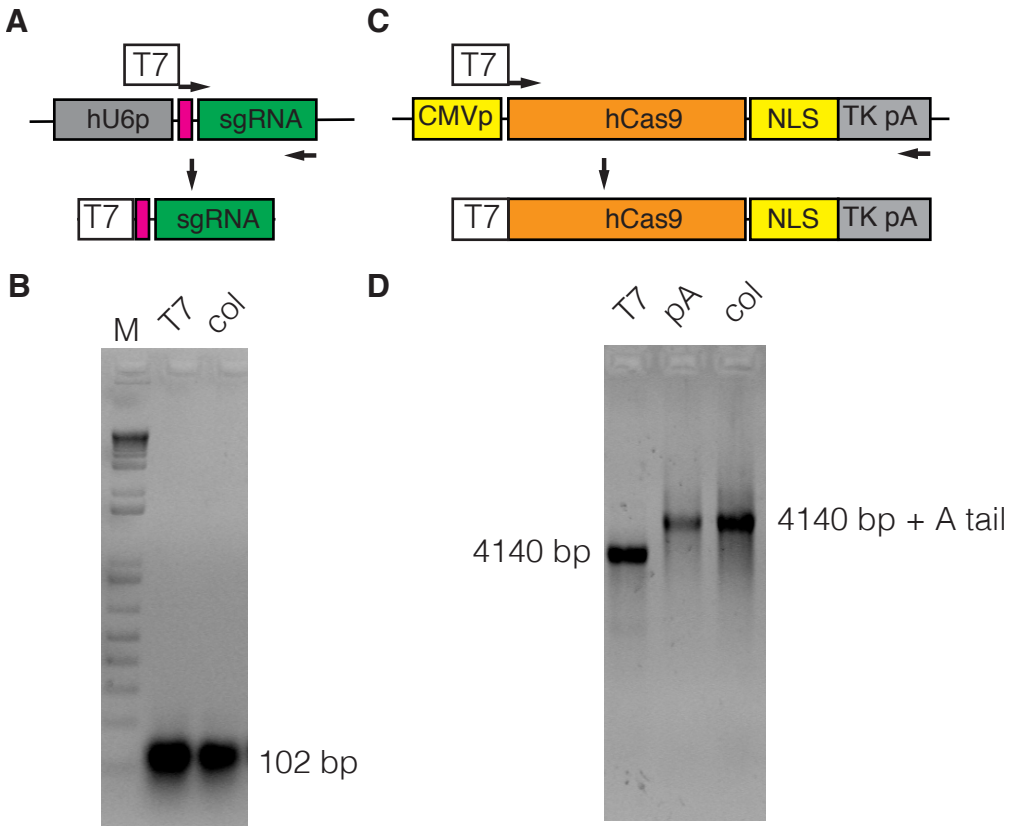


Figure m18 sgRNA and hCas9 *in vitro* transcription. **A** The sgRNA fragment is amplified with primers carrying the T7 RNA polymerase promoter. **B** The PCR product is used as template for *in vitro* transcription. The transcribed RNA is then purified using NucAway spin columns (Ambion). **C** The T7 promoter sequence is added to the hCas9 ORF by PCR. **D** One microgram of the purified PCR product is used for RNA *in vitro* transcription, followed by polyA-tailing. A clear supershift is observed after the A-tailing reaction.

The capped and A-tailed Cas9 mRNA was then purified using Nucaway Spin Columns (Ambion) rehydrated in microinjection buffer and stored at -80°C in 10 μl aliquots until necessary. The sgRNA region (a 20 bp target-specific sequence followed by 80 invariable nucleotides, shown if **Figure m19**) is PCR-amplified from MLM3636-based vectors (described in section 3.1.9). As in the case of the hCas9 mRNA, the T7 RNA polymerase promoter was added 5' to the Forward oligonucleotide. For this PCR, an invariable reverse oligonucleotide was combined with a forward oligonucleotide specific for the sgRNA to be amplified and carrying the T7 RNA promoter sequence (**Table m2**).

Table m2 PCR primers used to add the T7 RNA polymerase promoter to the sgRNA sequence.

```
sgRNA-FwJae 5'-TTAATACGACTCACTATAGGNNNNNNNNNNNNNNNNNNNN-3'
sgRNA-RvJae 5'-AAAAGCACCGACTCGGTGCC-3'
```

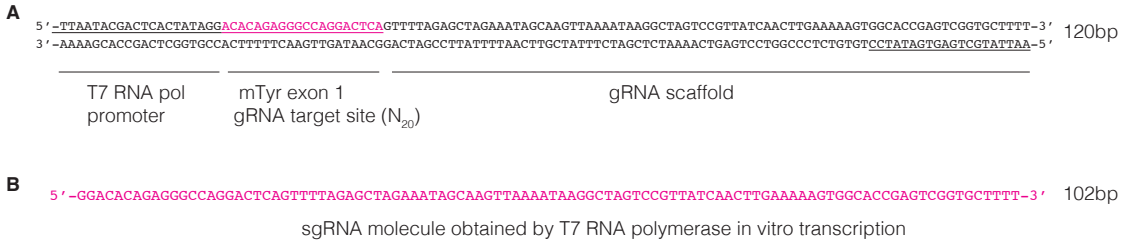


Figure m19 PCR template for sgRNA in vitro transcription. An example is given using a sgRNA targeting the mouse Tyrosinase exon 1. **A** Each target-specific sgRNA is PCR-amplified from MLM3636-based vectors using a Forward oligo carrying the T7 promoter sequence and a costant Reverse oligonucleotide (underlined). **B** The resulting RNA molecule is shown.

First, 10 ng of sgRNA plasmid were used for PCR with the following conditions:

1. +94°C	4 min
{2. +94°C	30 sec
3. +50°C	30 sec
4. +72°C	20 sec } 35 cycles
5. +72°C	2 min

Next, 500 ng of the sgRNA PCR was used for standard T7 RNA polymerase transcription, as follows:

T7 RNA polymerase 10x buffer	4 μl
10x BSA solution	4 μl
sgRNA PCR	5 μl
80mM rNTP solution (NEB)	1 μl
T7 RNA polymerase (NEB)	2 μl
H_20	up to 40 μl.

After o/n incubation at +37°C, the RNA was purified with Nucaway Spin Columns (Ambion) as done previously. RNA is stored in 10 μl aliquots at -80°C until necessary. Before proceeding with microinjection into mouse fertilized eggs, Cas9 mRNA and sgRNA RNAs are mixed and diluted to the proper working concentration in microinjection buffer (1 mM TrisHCl, 0.1 mM EDTA) and centrifuged at 30'000 x g for 1 hour at +4°C. These procedures have been recently described in details in Harms et al. 2014.

3.7 Transgenic mice

All procedures involving the use of mice complied with the European Union animal research legislation (2010/63/UE), as well as with the Spanish legislation (RD53/2013 and L32/2007, L6/2013). Also, they were reviewed and approved by the corresponding CNB and CSIC Ethics Committees on animal experimentation

in Spain or by the corresponding bodies in Switzerland, depending on the place of injection. Transgenic mice were produced at the Transgenic and Reproductive Techniques laboratory at the Institute for Animal Science, University of Zurich, Switzerland, headed by Pawel Pelczar, and by the Transgenic Core Facility of the CNB-CBMSO-CSIC, Madrid headed by Belen Pintado. RNA was microinjected into the cytoplasm (Horii et al. 2014) or into cytoplasm and pronucleus (to confirm positive microinjection) using standard procedures (Montoliu, 1997; Behringer et al. 2014; Yang et al. 2014; Harms, Quadros, Seruggia et al. 2014).

3.7.1 Genotyping

Mouse genomic DNA was extracted from tail or toe biopsies (< 1 cm) obtained at the time of weaning, in a two-step procedure (Montoliu, 1997). First, the tissue was digested overnight with a Proteinase K-containing lysis solution (0.5 mg/ml proteinase K, Roche; 100 mM NaCl, 50 nM Tris-HCl pH 8, 100 mM EDTA pH 8, 1% SDS, Merck). In the second step, the DNA was purified by ethanol precipitation, resuspended in TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8, Merck) and stored at 4ºC until further use. Alternatively, toe clips were digested overnight in 100 μl of a SDS-free lysis buffer (50 mM KCl, 10 mM TrisHCl pH 8, 0.45% NP40, 0.45% Tween-20, 500 ng/mL Proteinase K). The samples are then incubated 10' at 95C to inactivate the Proteinase K and used directly for PCR analyses. Transgenic founders were further analysed by sequencing. Briefly, 2 μl l of PCR were cloned into pGEM T-Easy vector (Promega) as indicated by the manufacturers. Several independent clones were sequenced using the external, universal oligonucleotides T7 or SP6.

3.7.2 Gene expression analyses in mouse tissues

Mice were euthanized by cervical dislocation. Organs were collected and snap-frozen in liquid nitrogen. RNAeasy kit (QIAGEN) was used to extract RNA as indicated by the manufacturer. Two μg of RNA were treated with RNAse-free DNAse (Roche). Subsequently, cDNA was obtained from 1 μg of RNA using Superscript III (Life technologies) and randon hexamer oligonucleotides (Life Technologies) as indicated by the vendor. One μg of RNA was kept without reverse transcription as a control for DNA contamination. A 1:20 dilution of cDNA

was used for RT-qPCR using gene-specific TaqMan Gene Expression Assays (Applied Biosystems) in a ABI 7500 machine. qPCRs were conducted in triplicates and repeated twice. Mm00495817_m1 probe was used for *Tyr*, Mm00690332_m1 for *Grm5* and Mm00479246_m1 for *Nox4* expression. Mm00446973_m1 probe was used for mouse TBP (TATA binding protein) expression, for normalization (Gimenez et al. 2003).

4. Results

4.1 The *Tyr* gene is flanked by chromatin boundaries

4.1.1 Synteny and gene expression analysis of the mouse *Tyr* locus

Synteny analyses (**Figure r1 A** and **Figure r2**) of the *Tyr* locus were performed by aligning known genomes from different vertebrates. Results indicate that evolutionary forces kept *Tyr* linked to its neighbouring genes, namely *Nox4* (Cheng et al. 2001), a nuclear-encoded mitochondrial protein involved in the respiratory chain, and *Grm5*, Glutamate metabotropic receptor 5, involved in the regulation of neural network activity and synaptic plasticity (Winder et al. 1996). Notably, these genes are not part of a gene cluster – a genomic structure often associated with syntenic conservation (Elizondo et al. 2009). On the contrary, they have dissimilar expression patterns (**Figure r1 B** and **Figure r3**). *Nox4* is broadly expressed in kidney, liver, and lung. *Grm5* is specifically expressed in several central nervous system structures including dorsal striatum, nucleus accumbens, cortex, hypothalamus, hippocampus as well as in peripheral nervous structures, such as the spinal cord and the neuroretina. *Tyr*, located in between *Nox4* and *Grm5*, is specifically expressed only in pigmented cells: in neural crest-derived skin and iris melanocytes and in the pigmented epithelium cells of the retina (Gimenez et al. 2003).

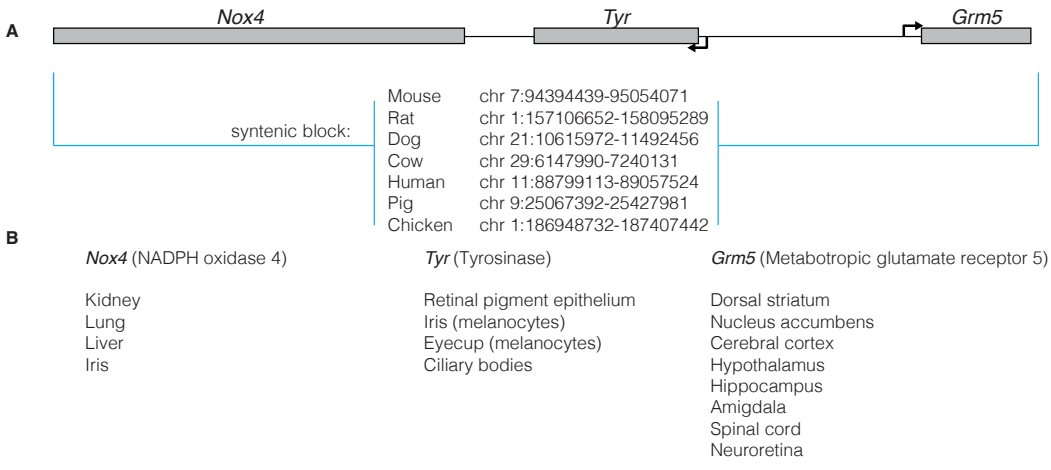


Figure r1: Differently expressed gene within the *Tyr* syntenic block **A** Synteny of the mouse *Tyr* locus was analysed using Ensembl. The *Tyr* locus is contained in a large syntenic block, conserved from chicken to human, along with the flanking genes *Nox4* and *Grm5*. A list of the tissues where each gene is expressed in mice – extracted from BIOGPS datasets – is shown in **B**. Note that the three genes are never coexpressed.

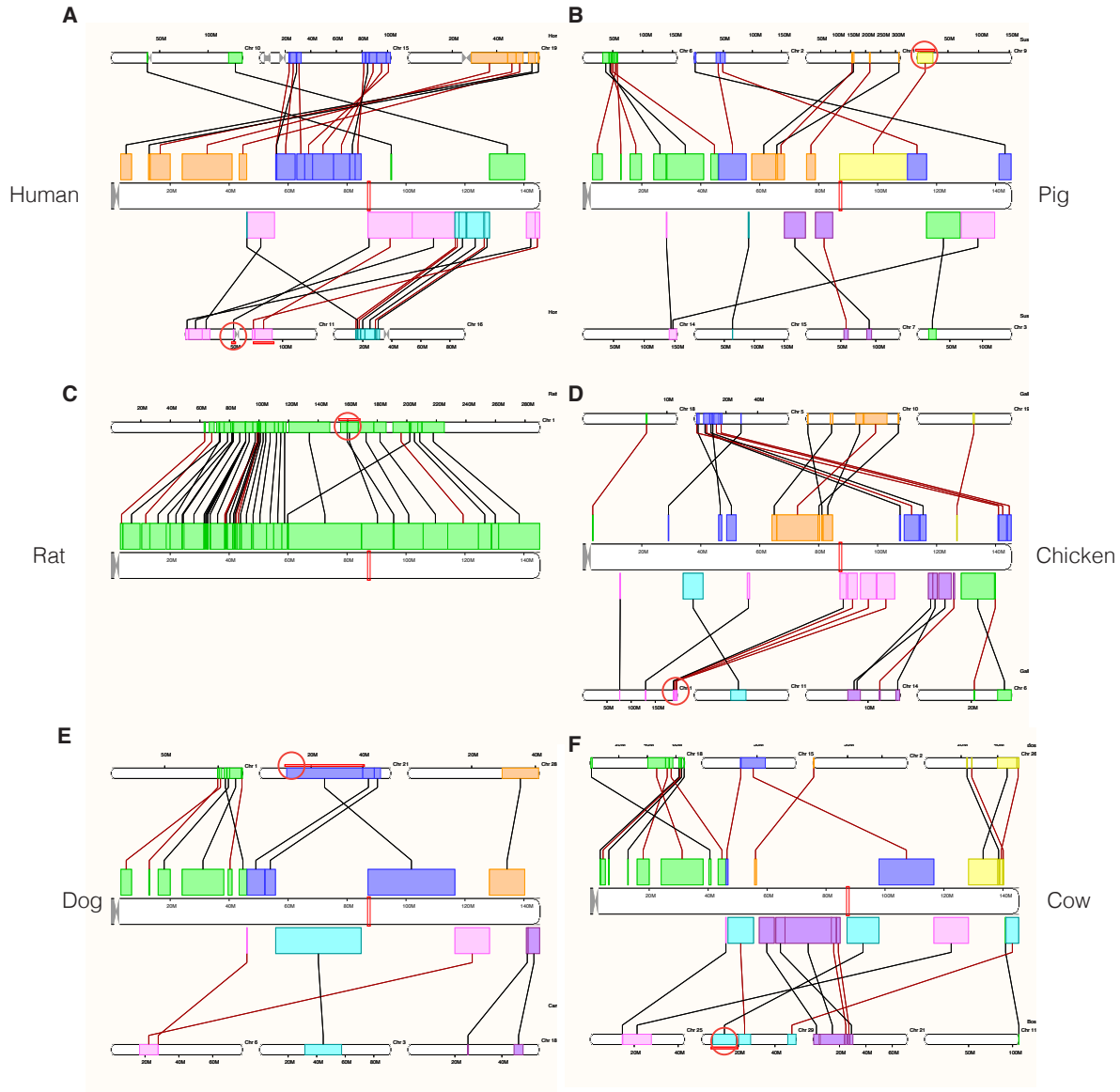


Figure r2: Synteny analyses of the Tyr locus across vertebrates (Ensembl). A detailed view of the synteny analyses obtained from Ensembl is given for all tested species. A region of the mouse chromosome 7, containing *Nox4*, *Tyr* and *Grm5* is syntenic with human chromosome 11 (**A**), pig chromosome 9 (**B**), rat chromosome 1(**C**), chicken chromsome 1 (**D**), dog chromosome 21 (**E**) and cow chromosome 29 (**F**).

The observation that these genes were kept physically together through evolution despite their divergent expression patterns let us hypothesize that chromatin boundaries – or insulators - must be present to insulate the *cis*-acting, regulating elements of the *Tyr* locus and preventing deleterious transcriptional cross-talk (Molto et al. 2009). The conserved synteny of *Nox4*, *Tyr* and *Grm5* indeed suggests that a fine mechanism have emerged, and eventually fixed and conserved, in order to protect the three independent expression units present in the locus.

Previously, 15 kb 5' upstream the *Tyr* gene, a chromatin boundary was identified and characterized extensively in our laboratory (Montoliu et al. 1996; Giraldo et al. 2003). The 5' *Tyr* boundary, or Locus Control Region (LCR), introduced in section 1.1.2 and 1.2.1, is the element that could contribute to the correct functioning and conservation of *Tyr-Grm5* syntenic block, and was also found in the human genome (Regales et al. 2003). The search and characterization of a potential *Tyr* 3' boundary, hypothetically needed to insulate *Tyr* and *Nox4* expression domains, is one of the objectives of this work.

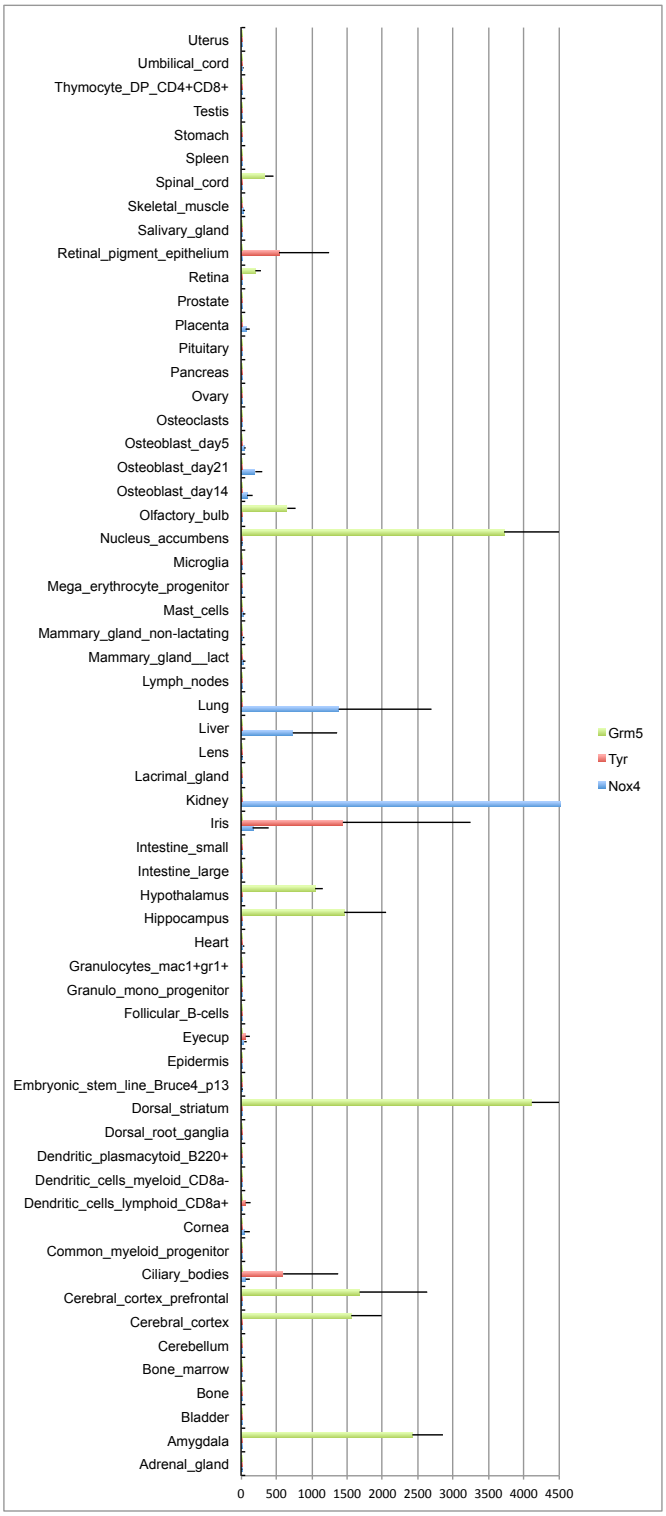


Figure r3: Expression profile of *Nox4*, *Tyr* and *Grm5* (BioGPS). Data were extracted from two independent datasets (GeneAtlas MOE430 1448821_at and 1417717_a_at). Means and standard deviations are shown for the three genes.

4.1.2 Identification of a *Tyr* 3' candidate insulator element

Preliminary experiments carried out in our laboratory suggested the existence of some CTCF binding sites within the *Tyr* 3' region (Angel Diaz, PhD thesis 2007). In order to identify new putative insulators 3' upstream the *Tyr* gene, we analysed the intergenic region between *Tyr* and *Nox4*. We used multi-species sequence conservation and we took advantage of several published CTCF and cohesin ChIP-seq datasets (**Figure r4**). In fact, cohesin contributes to the insulator activity mediated by CTCF (Wendt et al. 2008; Parelho et al. 2008) and to the formation of long-range contacts between distal DNA sequences, which is another mechanism to mediate insulator property. In addition, the presence of repetitive sequences and ancient transposable elements was also considered, since insulator activity was previously found associated with such elements (Lunyak et al. 2007; Roman et al. 2011). We focused our attention of a bi-partite 2.5 kb conserved non-coding element centered on the CTCF-Cohesin binding site. This non-coding region contains a CTCF-binding site, overlapping with a repeated element of the Short Interspersed element class B2 (SINEB2) family, a combination that often functionally serves as chromatin insulator. Odom and colleagues observed that among the all CTCF-binding sites found in several mammals, those that are embedded into mobile retro-elements are species-specific and often convey insulator activity (Schmidt et al. 2012). The likely origin of this type of boundaries depends on the internalization of a CTCF-binding site into an active retrotransposon, and in successive mobilization of the new retro-element carrying the CTCF binding motif. Such a mechanism is shared with many other transcription factor-binding sites, for example *Myc*, indicating that transposable elements play an important role in expanding the repertoire of their binding sites (Bourque et al. 2008). Those CTCF/transpositions events that result useful (or not detrimental) will be positively selected and fixed during evolution. Thus, an internal 241 bp fragment of the 2.5 kb conserved non-coding element, containing the strongest CTCF-binding ChIP-seq signal **(Figure r5 A**) and including the full SINEB2 element (**Figure r5B**), was also considered for functional validation.

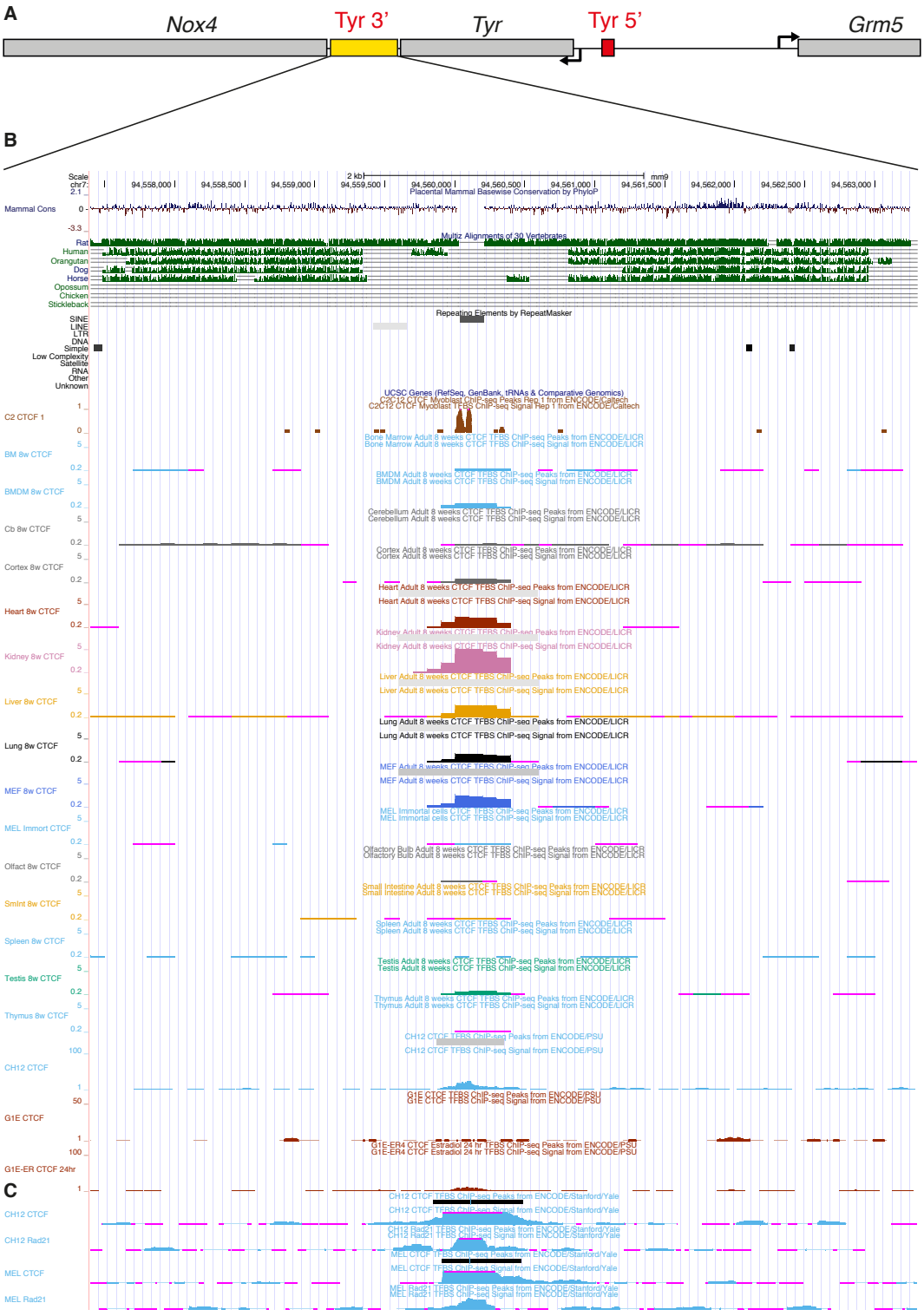


Figure r4 Identification of a *Tyr* 3' candidate insulator sequence. **A** Diagram of the *Tyr* locus. The *Tyr* 5' boundary is indicated as a red box. The *Grm5-Tyr* intergenic region is indicated as a yellow box. **B** Diagram extracted from UCSC Genome Browser illustrating mammalian, multi-species conservation of the *Grm5-Tyr* intergenic region, repeat elements and CTCF occupancy in several mouse tissues as found in several publicly available ChIP-seq datasets. **C** Colocalization of CTCF and Rad21-Cohesin binding.

Using the CTCF motifs dataset provided by Xie and co-workers (Xie et al. 2007), we identified a sequence resembling the CTCF LM7 motif in the selected putative *Tyr* 3' core element (**Figure r5 C** and **D**). By site-directed mutagenesis (see section 3.1.7 for detailed procedure), we introduced three point mutations (**Figure r5 E**) aimed to avoid or reduce the CTCF occupancy at the described sequence.

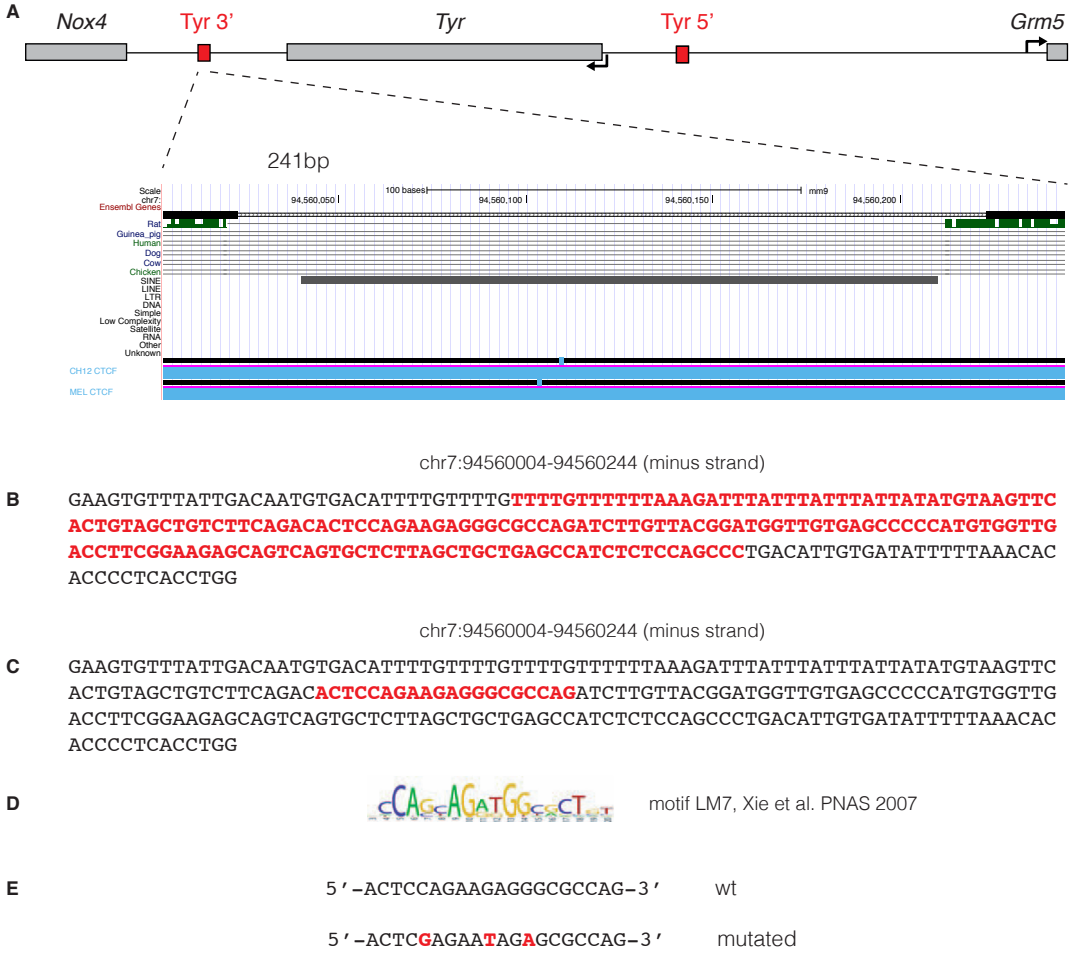


Figure r5 Sequence analysis of the putative Tyr 3' boundary. **A** The highest CTCF-binding signal is detected in correspondence of a 241 bp sequence, containing a SINEB2-type repeat (shown in red in **B**). **C** A CTCF-binding site was found matching the LM7 CTCF-binding motif, shown in **D**, described by Xie et al. 2007. **E** Three base pair in the CTCF-binding motif were mutated to eventually abolish CTCF binding.

4.1.3 Enhancer-blocking Assay in HEK 293 cells

We used the Enhancer-blocking assay in HEK 293 cells (Lunyak et al. 2007 and section 3.3) to confirm the insulator property of the previously described 5' boundary and to analyse the putative boundary on the 3' extremity of the *Tyr* locus.

Briefly, the candidate insulator sequences were cloned in a plasmid carrying the CMV enhancer and its minimal promoter, linked to the luciferase coding sequence (pELuc, section 3.1.1).

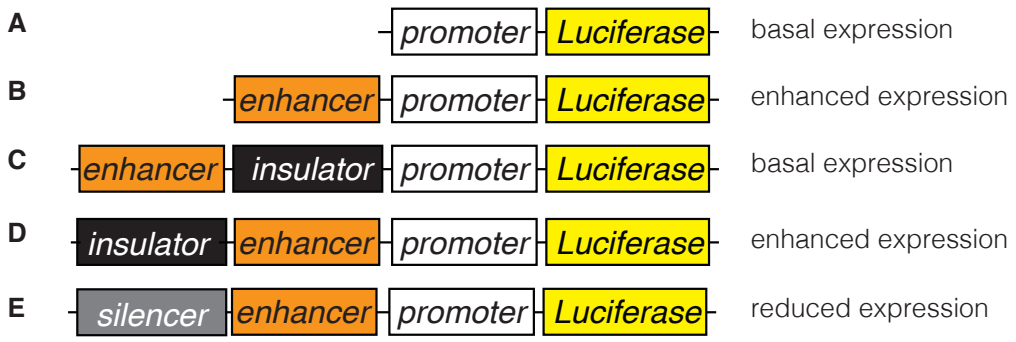


Figure r6: Principles of the Enhancer-Blocking Assay. Several luciferase-reporter constructs are transfected in HEK293 cells. A basal level of expression is observed when the minimal CMV promoter drives luciferase expression (**A**), while adding an enhancer will result in a higher luciferase activity (**B**). If an insulator is positioned in between the enhancer and the promoter, the luciferase expression drops down to a level similar to the basal vector (**C**), according to the strength of the insulator. When cloned upstream the enhancer, a true insulator does not affect reporter expression (**D**), while a silencer does (**E**).

Two series of vectors were generated. By cloning into the *Xho*I site of pELuc, the putative boundary is positioned in between the CMV enhancer and promoter. The reporter gene expression will sense the capability of the insert of blocking the enhancer activity: strong insulators will reduce the luciferase activity, whereas neutral sequences will not affect the reporter gene expression levels (**Figure r6**). A second series of vector was generated by cloning the candidate insulators in the *Pst*I site, which is located upstream the CMV enhancer. This series of vectors is meant to sense silencing activity of the cloned sequence. A reduction in the expression of the reporter gene in this series of vectors indicates that the cloned sequence is indeed not an insulator, but a silencer. Neutral sequences will not affect the reporter gene expression when cloned in this position. Along with test sequences, a set of control vectors is used: the cHS4 element (obtained from plasmid pBC1, described in Lunyak et al. 2007), from the chicken β-globin locus, and a smaller fragment of the latter (fragment II/III, Recillas-Targa et al. 1999; Recillas-Targa et al. 2002) were used as positive controls. A mutated version of the II/III fragment was used as negative control (Recillas-Targa et al. 1999).

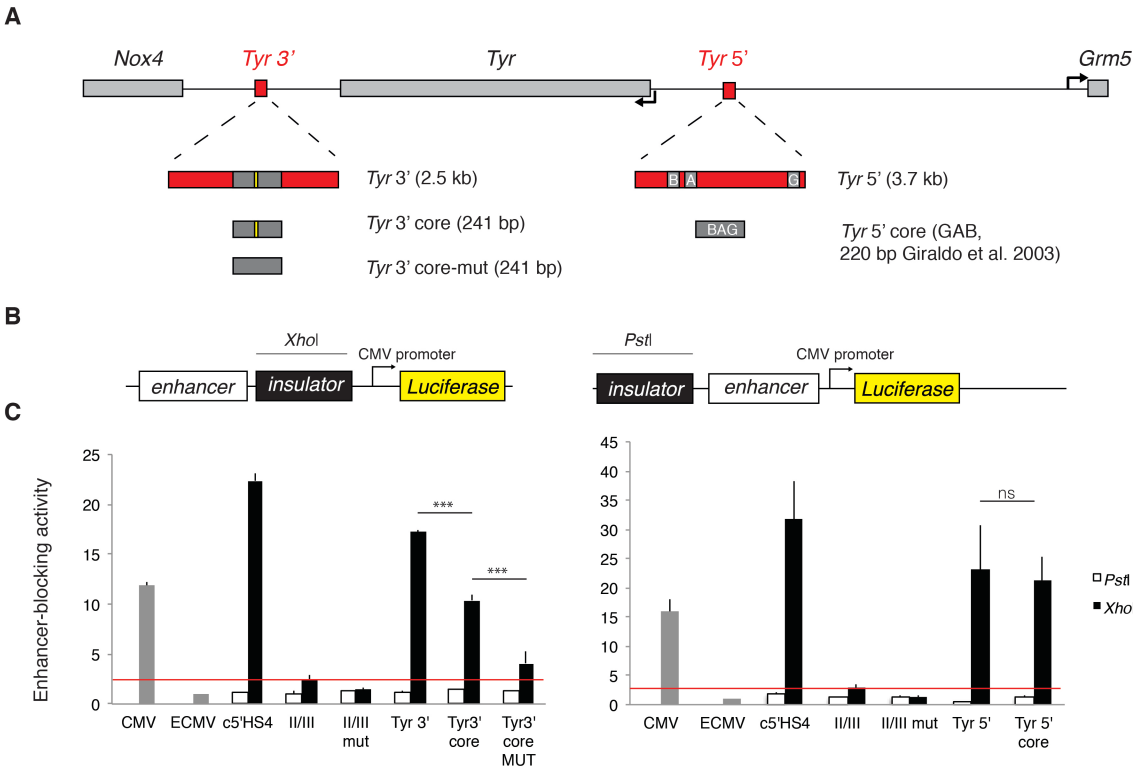


Figure r7 Enhancer-blocking assay in HEK 293 cells. **A** Candidate *Tyr* 3' boundary sequences, including a 2.5 kb fragment, a 241 bp subfragment and a mutated version were cloned into pELuc vector. The full 3.7 kb fragment of the *Tyr* 5' boundary (LCR, Giraldo et al. 2003) and the LCR core element GAB were cloned also in pELuc vector (**B**), downstream and upstream the CMV enhancer. **C** Enhancer-blocking activity is shown for the newly described *Tyr* 3' boundary and for the *Tyr* 5' boundary (LCR and GAB core), described in Giraldo et al. 2003. Anova-Bonferroni multiple comparison test. Significant $P<0.05$.

The candidate *Tyr* 3' insulator sequence that we identified has indeed insulator properties: the full-length, 2.5 kb *Tyr* 3' element displayed the highest enhancer-blocking activity. The *Tyr* 3' core element, a 241 bp element containing a CTCF-binding site embedded in a SINEB2 repeat, retains more than 50 % of the activity of the full-length counterpart in less than one tenth of the size. By mutagenesis of the CTCF-binding motif (**Figure r5 E**) the insulator activity is reduced, but not fully abolished. This result indicates that either the mutated CTCF-binding motif can still be partially bound by CTCF, or that the SINEB2 alone display a moderate boundary activity. EBA assay of the 3.7 kb *Tyr* 5' element (LCR) and its core element (GAB), confirms that this sequence, previously characterized in our laboratory (Giraldo et al. 2003), acts as a chromatin boundary (**Figure r7**). In conclusion, these results indicate that the *Tyr* expression unit is flanked by two elements capable of blocking enhancer activity *in vitro.*

4.1.4 Enhancer-blocking Assay in zebrafish embryos

Mammalian cultured cells represent a useful model to easily test several elements in parallel, in a relatively fast and cheap way. Nevertheless, *in vitro* assays lack the higher level of complexity, typical of a living organism. To overcome this limitation, we developed in the laboratory, and in collaboration with JL Gomez-Skarmeta (CABD, Seville) a second series of Enhancer-Blocking Assays, using zebrafish embryos as a model organism (Bessa et al. 2009; Roman et al. 2011; Martin et al. 2011; Tiana et al. 2012), described in section 3.3.

Using the p48RCAR vector (section 3.1.2) the putative insulator is positioned in between a midbrain-specific enhancer and a muscle-specific promoter. Thanks to the Tol2-transposase system (Kawakami et al. 2007), transient-transgenic individuals are generated with high efficiency by embryo microinjection. Injected embryos are imaged by fluorescence microscopy 36 hours post-fertilization. The ratio of GFP fluorescence between CNS and somites will be indicative of the level of insulator activity, low values indicating neutral/weak sequences and high values indicating sequences capable of blocking enhancers.

We used this assay to confirm the properties of the *Tyr* 3' core element *in vivo*. We included in the experiment the mutated *Tyr* 3' core element, to confirm that the enhancer-blocking activity of the *Tyr* 3' boundary is largely CTCF-dependent.

The *in vivo* enhancer-blocking assay confirmed the insulator property of the *Tyr* 3' core element (**Figure r8 C** and **E**), which shows a statistically significant reduction in mid-brain fluorescence ($p < 0.000726$). When a mutated *Tyr* 3' element, carrying three mutation in the CTCF-binding site, the enhancer-blocking activity is reduced (**Figure r8 F**) and no statistically significant difference is observed when compared with control embryos ($p < 0.135$; **Figure r8 C** and **D**), thereby proving that the boundary activity has been impaired *in vivo.*

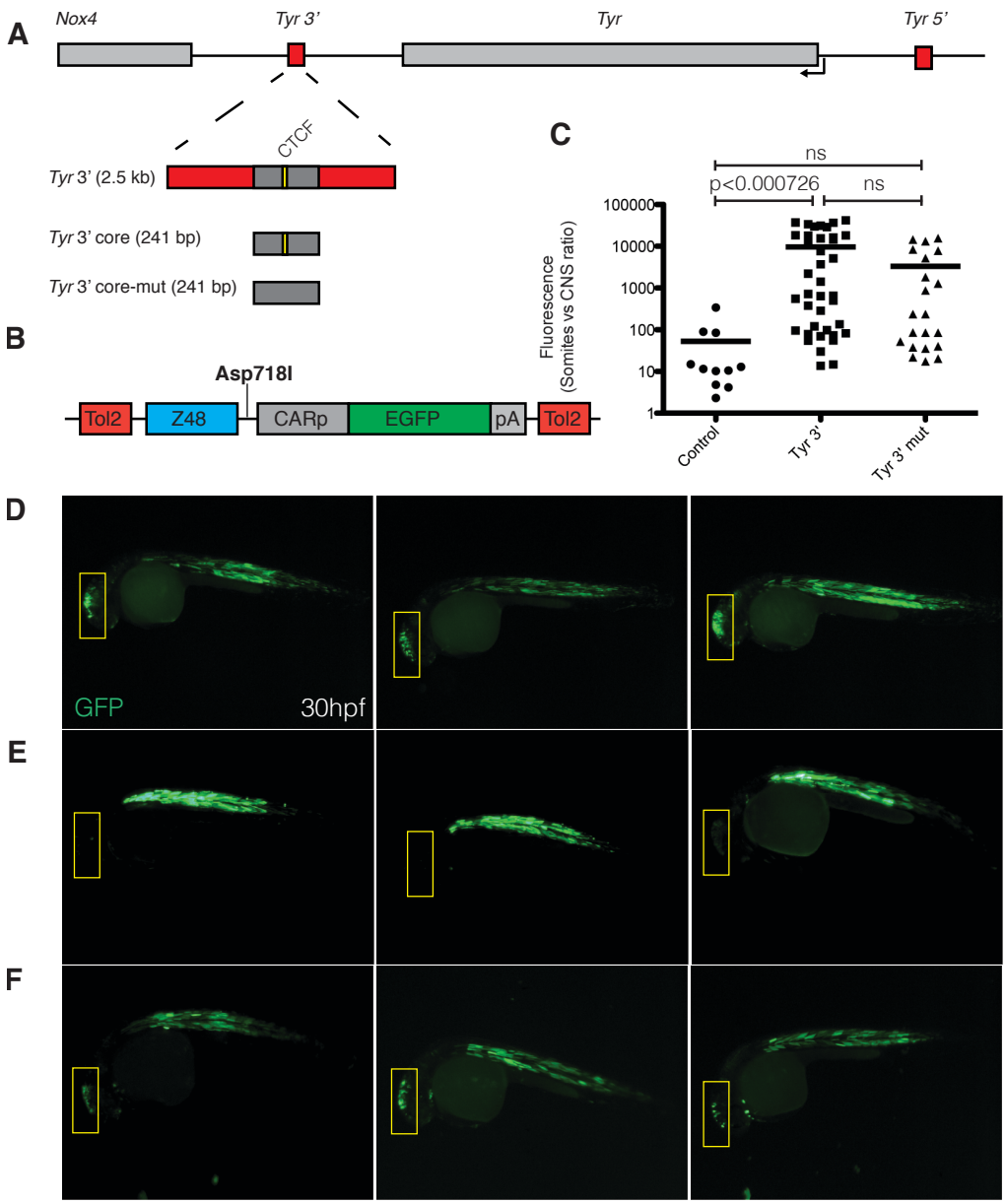


Figure r8 Enhancer-blocking assay in zebrafish embryos. **A** 241 bp of the *Tyr* 5' core boundary, and its CTCF-mutated counterpart were cloned in the pCAR48R plasmid. **B** This vector contains a midbrain-specific enhancer and a muscle-specific minimal promoter, linked to the GFP coding sequence. This expression cassette is flanked by Tol2 inverted sites to be used for transposase-mediated transgenesis in zebrafish embryos. **C** Ratio of somites vs brain fluorescence. Differences between groups were examined for statistical significance using Median test; n = 11-39. *Tyr* 3' versus control, p<0.00726; *Tyr* 3' mut versus control, p<0.135; *Tyr* 3' versus *Tyr* 3' mut, p<0.591. **D** Control embryos injected with the empty pCAR48R vector. Expression in both midbrain and somites can be observed **E** Representative embryos injected with the *Tyr* 3' construct. **F** Embryos injected with the mutated *Tyr* 3' element. The location of midbrain is indicated by yellow boxes.

4.2 Tri-dimensional organization of the mouse *Tyr* locus

In the previous section we focused on the enhancer-blocking property associated with the *Tyr* 5' and 3' insulators. However, chromatin insulators and other *cis*-acting elements play additional roles in the genome organization (West et al. 2002; Molto et al. 2009). Insulators are able to *physically* partition individual expression domains into independent chromatin loops (Yang and Corces, 2012; Dixon et al. 2012). Such structural role mainly rely on the capability of insulators to recruit CTCF, a transcription factor often referred as the ''main vertebrate insulator protein (Phillips and Corces, 2009). CTCF is able to form dimers (Williams and Flavell, 2008), putting in contact distal DNA region bound by CTCF (or other protein partners) resulting in the formation of a chromatin loop with the intervening DNA inside (Hou et al. 2008).

With this in mind, we aimed to study the physical structure of the *Tyr* locus. We used Chromosome Conformation Capture (3C, de Laat and Dekker, 2012) to observe the *Tyr* locus architecture in an active and inactive status, using appropriate mouse cell lines and experimental designs.

4.2.1 *Tyr* expression in B16 and L929 cells

B16-F1 mouse melanoma cells (Ganss et al. 1994) were chosen as a cellular model in which the *Tyr* gene is actively transcribed. We confirmed that B16 cell line expresses *Tyr* by mRNA quantification (**Figure r9 A**) and by the observation that melanin, the downstream product of the tyrosinase enzyme, is accumulated in these cells and clearly visible (**Figure r9 B**) in cellular pellets. On the other side, in mouse L929 fibroblast cells we could not detect *Tyr* mRNA or melanin. Therefore L929 cells were used to model a non-transcribing, inactive *Tyr* locus.

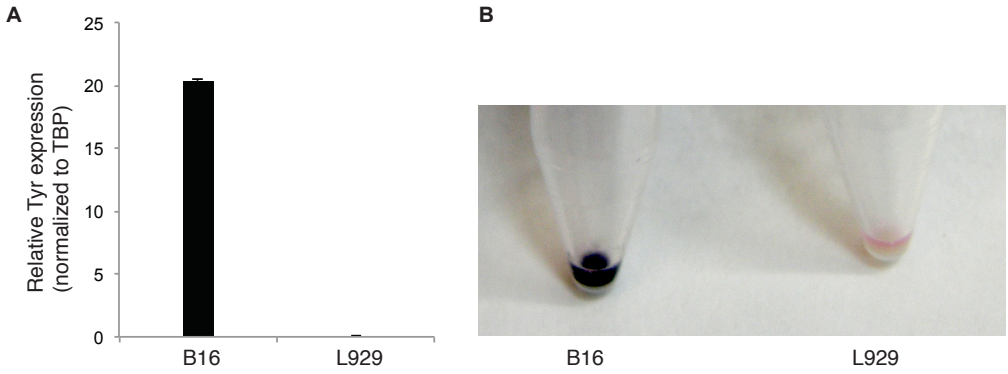
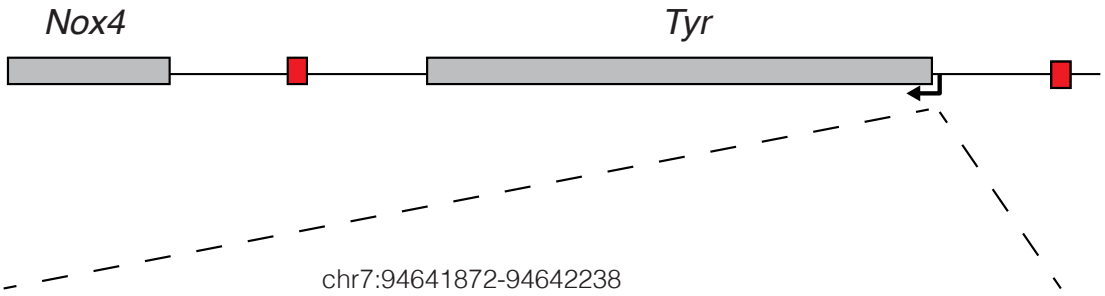


Figure r9: Tyr expression cultured cells. **A** Taqman RT real-time PCR quantification of *Tyr* mRNA in B16 and L929 cells. *Tyr* mRNA levels are normalized to that of TATA-binding protein mRNA levels. **B** Melanin, downstream product of the enzyme tyrosinase, is accumulated in B16 cells, but not in L929 cells.

4.2.2 Chromosome Conformation Capture (3C)

The *Tyr* promoter contains the minimal information necessary to correctly drive *Tyr* expression *in vivo* (Beerman et al. 1990; Kluppel et al. 1991; Tanaka et al. 1990; Yokoyama et al. 1990). This information is conveyed by two conserved motifs, the *Tyr Initiator* (*Inr*) and the M-box (Bentley et al. 1994), both bound by the transcription factor *Mitf*. We therefore selected a 367 bp *Dpn*II fragment containing these elements as anchor fragment for the 3C assay (**Figure r10**). Combining oligonucleotides mapping on several *Dpn*II fragments in the *Tyr* locus with a primer mapping on the promoter, we interrogated our 3C libraries (section 3.4) in order to detect which DNA elements were physically contacting the *Tyr* promoter in *Tyr*-positive and negative cells, with the highest resolution (Comet et al. 2011; Sexton et al. 2012).

Results of the 3C assay indicate that the *Tyr* promoter acts as the hub of chromatin interactions in the *Tyr* locus (**Figure r11 B**). A strong interaction is observed between the promoter and a fragment 80 kb 3' upstream (highlighted by circle number 1). This *Dpn*II fragment colocalizes with the *Tyr* 3' boundary element, previously characterized by the Enhancer-blocking assay in human cells (2.1.3) and zebrafish embryos (2.1.4). The interaction is observed in both B16 cells and L929, indicating the constitutive nature of the chromatin loop, likely having a structural role, in virtually all cell types, independently from *Tyr* expression. The *Tyr* 3' boundary sequence is in fact occupied by both Rad21 and CTCF. No interactions are observed with control sequences immediately upstream and downstream the *Tyr* 3' element.



GATCATACAAAATCTGCACCAATAGGTTAATGAGTGTCACAGACTTCTTTTTCCAGCAACCCCTGGAGT
GACTATCACATGTTTTGGCTAAGACCTATATAACCACTCCCCTTCTTACATCAAATACTCTCAGCCTG
TTTTACACTAAGCTTTTATCTTCTGCAAAGCACATGACTAACTTTTTCTTGGAGTTTGTACATAGCCC
ATAGTGAGGTAACTAAATTGAAGGAAGATATATTATTCTAATTGATATGAAATTAATAATAATTGGAA
TTTGCAAATGAGAAGTAGTATCAAGAATACATCATCTCTTATGAATAAAGGTCATGAAACATCTAGAT
TCTGAATACAAGCCTGTTGTATTTAGA

Figure r10 *Dpn*II fragment containing the *Tyr* promoter. A 367 bp *Dpn*II fragment was selected as anchor for the 3C study, containing the *Tyr* promoter Initiator (in red) and the M-box (in green). The sequence of the oligonucleotide used is underlined. Genomic sequences are extracted from the mouse genome assembly NCBI37/mm9 (July 2007).

A second interaction is observed 48 kb 3' upstream the *Tyr* promoter, specifically in the B16 cell 3C library, located in the 3rd *tyrosinase* intron (circle 2). Regulatory elements are often found in introns (i.e. Ried et al. 1990). Conversely, in the first intron, which was included in most *tyrosinase* minigene constructs (Beermann et al. 1990), there is no detectable chromatin interaction. 15 kb 5' upstream the *Tyr* TSS, a cell-type specific interaction is found with the promoter (circle 3). The interacting sequence co-localizes with the *Tyr* 5' element, or LCR. The interaction is restricted to a *Dpn*II fragment containing the *Tyr* LCR A and B boxes (Ganss et al. 1994; Giraldo et al. 2003). Flanking, control fragments are not interacting with the *Tyr* promoter. Notably, a *Dpn*II fragment containing the LCR G box, (Giraldo el al. 2003) is not interacting with the *Tyr* promoter (circle 4). No interaction is observed with far 5' upstream elements, including one element previously described to be involved in *Tyr* expression regulation (Murisier et al. 2007) (circle 5). Next, the 3C libraries were interrogated using the *Dpn*II fragment containing the *Tyr* 5' element as bait region. The *Tyr* locus was scanned for sequences interacting with the *Tyr* 5' element (**Figure r11 C**). Interactions are only observed with the *Tyr* promoter (circle 6) and not with the Grm5 promoter, located 75 kb 5' upstream, indicating that this regulatory element is not shared with other expression units. No interactions are observed between the *Tyr* 5' and *Tyr* 3'

elements, indicating that these two boundary elements probably compete with the same binding sites in the *Tyr* promoter.

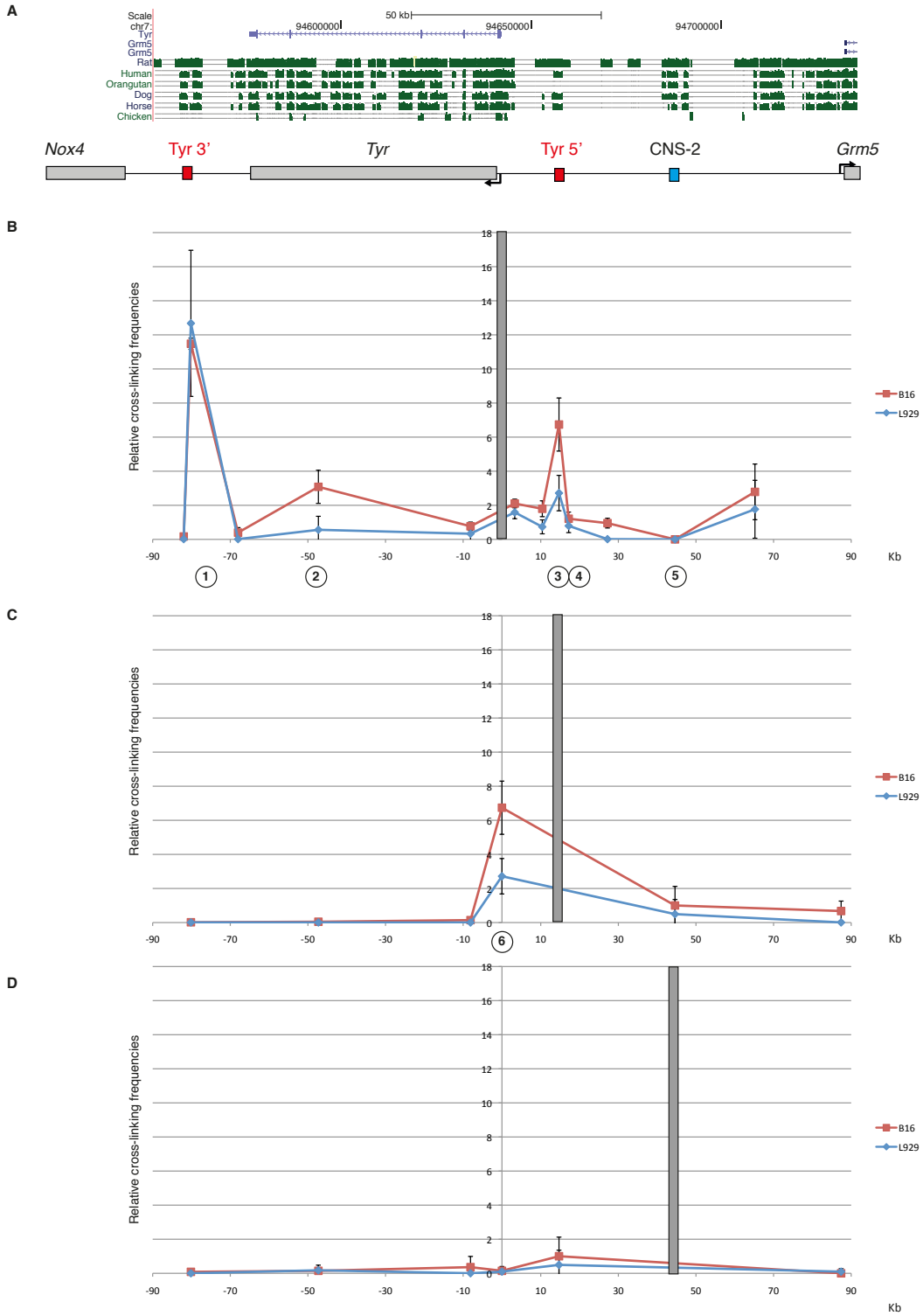


Figure r11 3C study of the Tyr locus. **A** Diagram of the *Tyr* locus along with multi-species sequence conservation track obtained from the UCSC Genome Browser. The *Tyr* 3' and 5' boundary elements are shown as red boxed. The CNS-2 sequence (Murisier et al. 2007) is shown as a blue box. **B** Chromosome conformation capture using a bait primer located on the *Tyr* promoter. **C** 3C using a bait on the *Tyr* 5' boundary fragment containining the AB boxes. **D** 3C using a bait on the CNS-2, - 47 kb element.

Finally, the anchor primer was positioned on a fragment containing a conserved region described by Murisier et al. 2007 and observed to be dispensable in transgenic mice generated with different YAC *Tyr* transgenes (Montoliu et al. 1996). Using this bait fragment, we could not observe any interaction within the *Tyr* locus in both 3C libraries (**Figure r11 D**) in the cell lines we used.

4.2.3 Sequencing confirmation

In order to confirm the 3C findings, the PCR products deriving from chimeric ligation products were cloned and sequenced using universal sequencing oligonucleotide contained in the cloning vector. **Figure 12** shows the sequencing confirmation of the interaction detected between the *Tyr* promoter and the *Tyr* 5' boundary element located 15 kb upstream. Alignments of the sequences of DNA strands using BLAT indicate that the PCR amplicon indeed represent a chimeric ligation product of two distal DNA fragments (**Figure r12 B**).

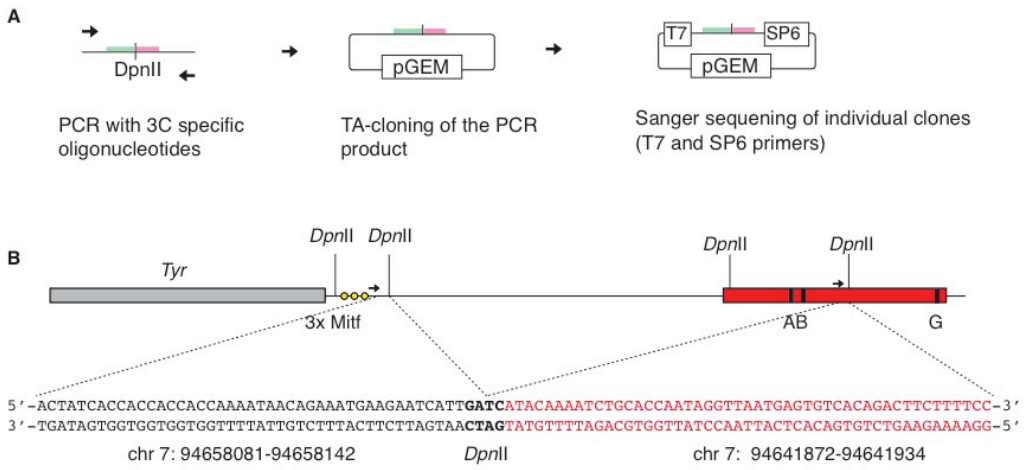


Figure r12: Sequencing confirmation of *Tyr* 5'-*Tyr* promoter interaction. **A** A chimeric PCR product obtained with oligonucleotides mapping on distal region in the *Tyr* locus was TA-cloned using pGEM T-Easy vector. Independent clones were sequenced using T7 and SP6 oligonucleotides. **B** Scheme of the position of the primers used to generate the chimeric fragment that has been sequenced. Genomic coordinates refer to the mouse genome assembly NCBI37/mm9 (July 2007).

Figure r13 illustrates the sequencing confirmation of the interaction detected between the *Tyr* promoter and the *Tyr* 3' element located 80 kb upstream. Sequence alignments of both DNA strands using BLAT indicate that the PCR product indeed represent a chimeric ligation product of two distal DNA fragments.

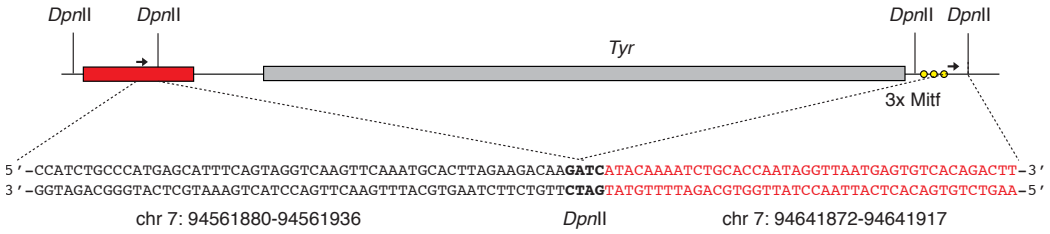


Figure r13 Sequencing confirmation of *Tyr* 3'-*Tyr* promoter interaction. A chimeric PCR product obtained with oligonucleotides mapping on distal region in the *Tyr* locus was TA-cloned using pGEM T-Easy vector. T7 and SP6 sequencing confirmed that the PCR product contains a chimeric sequence derived from the *Tyr* 3' boundary (red box) and *Tyr* promoter *Dpn*II fragments. Genomic coordinates refer to the mouse genome assembly NCBI37/mm9 (July 2007).

4.2.4 Structure of the *Tyr* locus in mouse RPE cells

As introduced elsewhere in this work, *Tyr* is expressed in neural crest-derived melanocytes and in RPE cells that derive from a different embryonic structure, the optic cup. To test whether different regulatory elements are required to activate *Tyr* expression in a different lineage, we wanted to perform 3C on RPE cells. Although several established human RPE cell lines are available, including ARPE-19 (Dunn et al. 1996) and D407 cells (Davis et al. 1995), only one mouse RPE cell line is available, called B6-RPE07 (Chen et al. 2008), a cell line cloned from a primary culture of mouse RPE cells. We obtained this cell line from Dr. Heping Xu (Centre for Vision and Vascular Science, Queen's University Belfast, UK). Unfortunately, B6-RPE07 cells do not express *Tyr*, nor the typical RPE-markers *Rpe65* and *Otx2* (**Figure r14 A**). Actually, loss of *Tyr* expression was readily detected also in primary cultures of mouse RPE cells (**Figure r14 A**), probably as a consequence of the Epithelial-mesenchymal transition (EMT) and likely caused by *Zeb1*, a inhibitory transcription factor that binds to E-box elements, also found at the *Tyr* and *Mitf* promoters. Liu and colleagues reported that the formation of spheres reverses EMT, represses *Zeb1* and restores both *Tyr* and *Mitf* expression (Liu et al. 2009). By obtained spheres from the B6-RPE07 cells (**Figure r14 B**), we observed an increase in *Mitf* expression but not in *Tyr* expression or a reduction in *Zeb1* levels. These technical limitations avoid, for the moment, the study of the structure of the *Tyr* locus in RPE cells, by 3C.

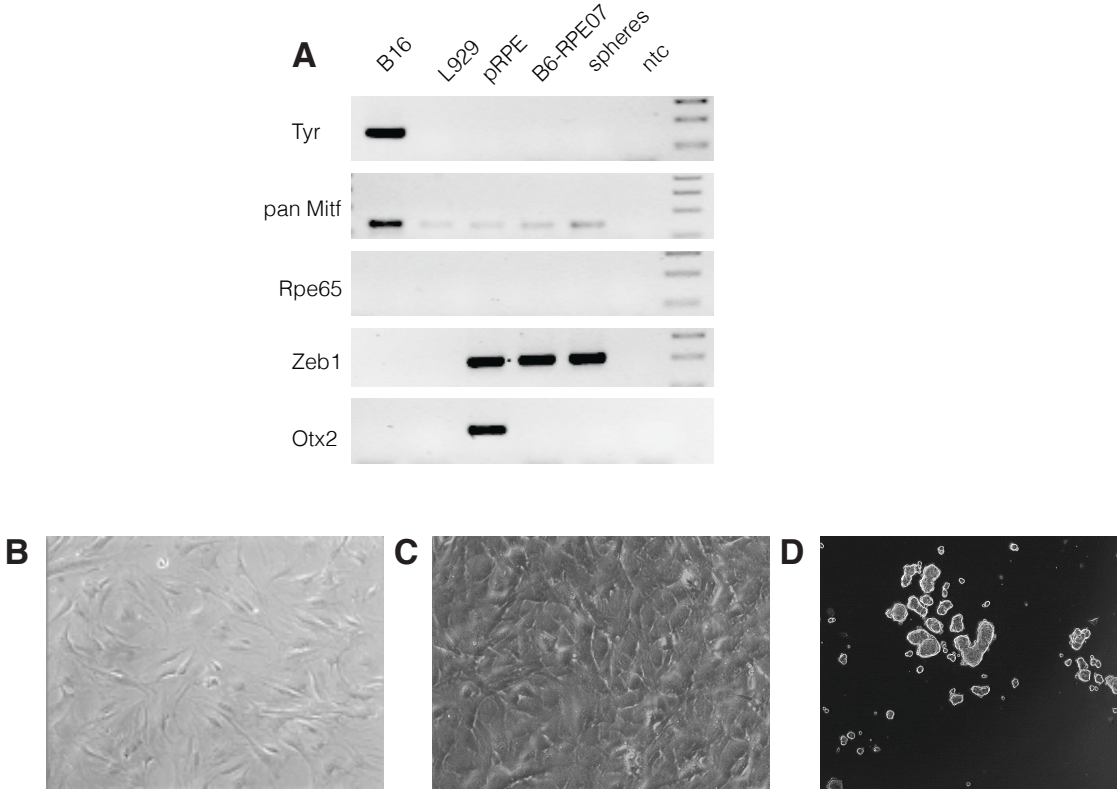


Figure r14 Characterization of the mouse B6-RPE07 cell line. **A** RT-PCR analysis. *Tyr* expression is detected in B16 mouse melanoma cells, but not in L929 fibroblasts, primary mouse RPE cultures at day 3, in B6-RPE07 and in sphere obtained for the latter. *Mitf* is strongly expressed in B16 cells and weakly in all other tested cells using primers detecting all *Mitf* isoforms. The formation of spheres increases *Mitf* expression. *Rpe65* cannot be detected in all tested cells. *Zeb1* is strongly expressed in primary RPE at day 3 of culture, in B6-RPE07 cells and in spheres. *Otx2* is only detected in primary RPE cells. **B** Morphology of primary mouse RPE cultures. **C** Morphology of the B6-RPE07 cell lines. **D** Spheres obtained by culturing B6-RPE07 cells in non-adherent plates.

4.3 Chromosomal inactivation of the mouse *Tyr* boundaries

In the previous sections the two identified boundaries in the mouse *Tyr* locus were studied using two different systems. First, the putative insulators were challenged using an artificial locus composed by a heterologous enhancer and promoter, linked to a reporter gene and aimed to assess the effect of the boundary in interfering the enhancer-promoter interaction, in cultured cells and in zebrafish embryos. Then, by Chromosome Conformation Capture (3C), the *Tyr* boundaries were analysed from the structural point of view, in their natural chromosomal context, *in vitro*, using cellular models mimicking an active (B16 cell line) and inactive (NIH 3T3 - L929 cell line) *Tyr* expression unit. By performing the 3C study, we gained insight into the importance of the interaction of boundaries and other *cis*-acting elements with respect of the *Tyr* promoter. However, this approach is still an approximation, considering that stable cell lines are only a

''snapshot'' of the real biological complexity, made of different tissues, cell types and developmental stages. To investigate the function and structure of the *Tyr* locus in a higher complexity, it is necessary to produce genetically modified mice harbouring the targeted inactivation of the two *Tyr* boundaries. Only with these animal models, the mechanisms that regulate the activity of *Nox4*, *Tyr* and *Grm5* – and the role of *Tyr* boundaries – can be fully described. In the next paragraphs, the generation of transgenic mouse lines bearing the inactivation of the *Tyr* 5' and *Tyr* 3' boundaries using targeting nucleases is presented.

4.3.1 Transcription Activator-Like Effector Nucleases (TALENs)

First, we used targeted nucleases (1.3.5) to achieve the chromosomal inactivation of the desired *cis*-acting elements in mice. We first focused on TALEN nucleases, introduced in section 1.3.5.2. Two TALEN series flanking the *Tyr* 3' and 5' boundaries were designed (**Table r1**) as described in section 3.5 using TALEN Targeter online tool (Doyle et al., 2012) and assembled by Golden Gate Cloning (3.1.9). Each TALEN DNA-binding array was cloned into two different series of expression vectors. The first series of TALEN vectors are based on homodimeric *Fok*I nuclease domain. The second series of TALEN vectors are based on an improved obligate heterodimeric *Fok*I domain (Guo et al. 2010; Doyon et al. 2011) (section 3.1.4). **Figure r15** illustrates the chromosomal position of the twelve TALEN pairs assembled for this study. Six TALEN pairs were assembled in our laboratory for the inactivation of the *Tyr* 5' boundary (**Figure r15 A**). Because of the presence of a large LINE1 repeat, the G-box element of the *Tyr* 5' boundary could not be targeted using TALENs. Another six pairs of TALENs were produced for the chromosomal inactivation of the *Tyr* 3' boundary (**Figure r15 B**).

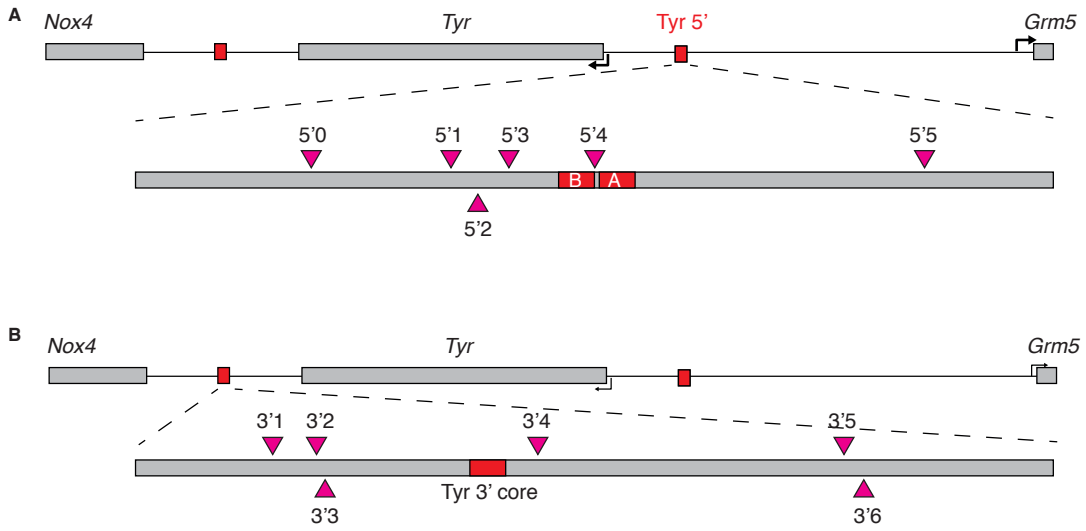


Figure r15 TALENs assembled in this study. **A** Six TALEN pairs were designed to target the *Tyr* 5' boundary. **B** Similarly, six TALENs were design to inactivate the *Tyr* 3' boundary. TALEN target sites are indicated as purple triangles. Core elements of each insulator are indicated as red boxes.

Table r1: TALEN target and spacer sequences. Left and right target sequence, and spacer sequences are provided for the TALENs generated in this work.

TALEN	target sequence	spacer	target sequence
5' 0	TACTATCTTTCACTGAATTGC	TCATGTTGCTTCCTGACCCTTGG	AGATGTCTGAATTGGAGGAG
5' 1	TATTTTCTTACCACCCACCC	TGTGTTATTGTGCTT	AAGGCTTTAGAGGAATATGA
5' 2	TCTTCATTCAGTAGACCAAAG	AAATGCATATTTTGTGTTTTTAT	TTTCCCAAGAATTAAGTGTGA
5' 3	TAAACAAGCCACATTCC	GCTTTGTAGAAGTTC	ACAATCTGAATAGGAATATAA
5' 4	TGGAACACAGACATCACTGTC	TTGACAAAGTCAGCTGGC	AGGAACTATGACCTTGGAGA
5' 5	TACCCAAAATTTAAATCTAAC	AGAGGATTGATGAGTCATTT	AGTCTAGAAAATCTGTAATAA
3' 1	TGAACACATTAACTAAAAAAA	GCAGTTTTTTTTTTAAATT	TTCAACTTGATCTTGAATGTA
3' 2	TGATTTGCAATTGCCATTACC	AAAAATTTTTATTTTC	TTGTGGTTATAATTTGTTTAA
3' 3	TGCCATTACCAAAAATTTTTA	TTTTCTTGTGGTTATAATTTGTT	TAAATCAATCTGGATTTAAGA
3' 4	TTAAAGTATCTATGCATCTCT	ACAGCTGATGGTAAA	AAAGAATGTTCGAATGAAAA
3' 5	TTCTTTTCACAACACTGTCAT	ACCATCTGTAATGAGCATTTC	AGTAGGTCAAGTTCAAATGC
3' 6	TACATTCTTCATCCACATTAT	CTGCTTTGTGTCAAA	ATGGGGTCATGAGTACATG

4.3.2 Deleting *Tyr* boundaries *in vitro* with TALENs

We performed the *in vitro* cleavage assay (adapted from Jinek et al. 2011, section 3.5.3) to validate the activity of each of the assembled TALENs. Two plasmids encoding the two monomers of each TALEN pair were transfected in HEK 293 cells. Cells were then lysed and protein extracts were used to digest a PCR product containing the target site of each nuclease. With this assay, we aimed to verify that each individual TALEN had been assembled correctly and that was able to cleave the corresponding target site. As shown in **Figure r16**, all

six TALEN pairs of the *Tyr* 5' series are able to cleave the respective target sequence, with exception of the *Tyr* 5'3 TALEN. Notably, the cleavage is more efficient when obligate-heterodimeric vectors (*Fok*I$^{ELD/KKR}$ mutant) were used. For this reason, obligate-heterodimeric vector series were used systematically in the following experiments.

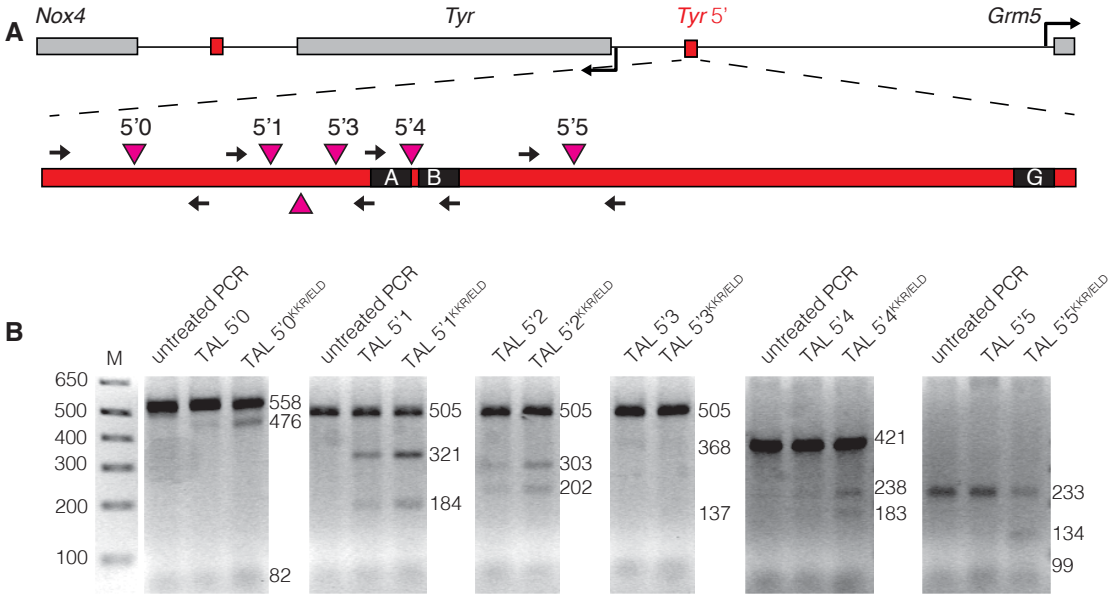


Figure r16 In vitro cleavage assay of the Tyr 5' TALENs. **A** Position of the TALEN target sites across the *Tyr* 5' sequence. The A, B and G boxes (Giraldo et al. 2003) are indicated as black boxes. TALENs binding sites are indicated as red triangles **B** Protein extracts of cells transfected with each TALENs pair were incubated with a PCR product containing the corresponding target sequence. Each TALE array was cloned and tested in the homodimeric *Fok*I vector and in its heterodimeric derivatives, the mutants *Fok*I$^{ELD/KKR}$.

The same *in vitro* cleavage assay was performed with the *Tyr* 3' series, in order to validate that TALENs were correctly assembled, and that the resulting artificial protein is able to recognize and cut its target sequence.

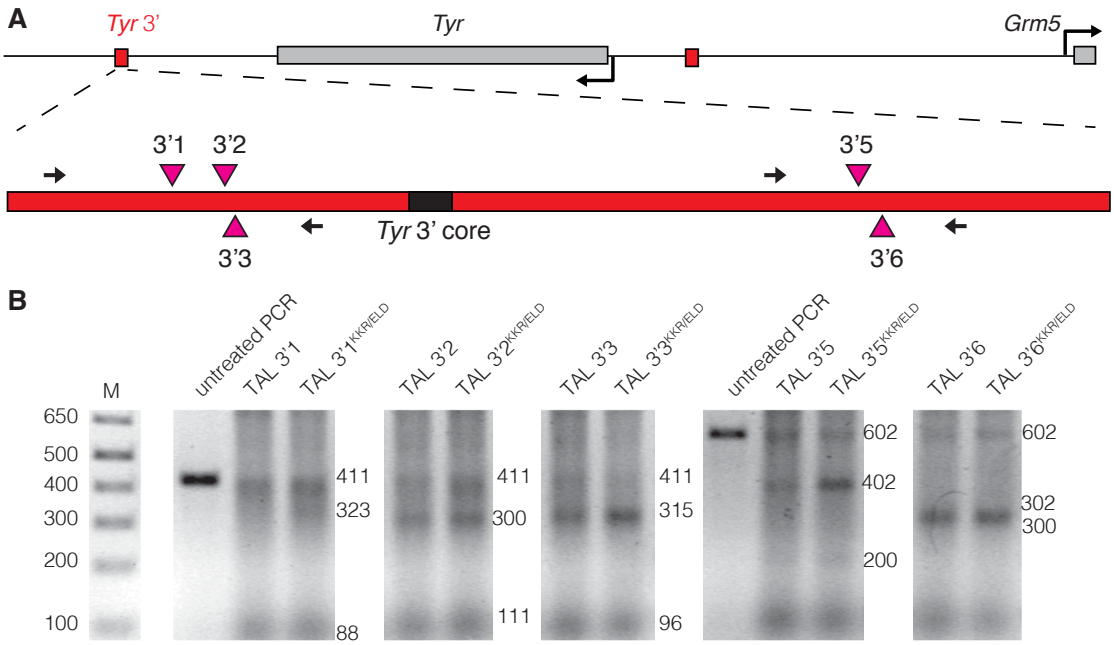
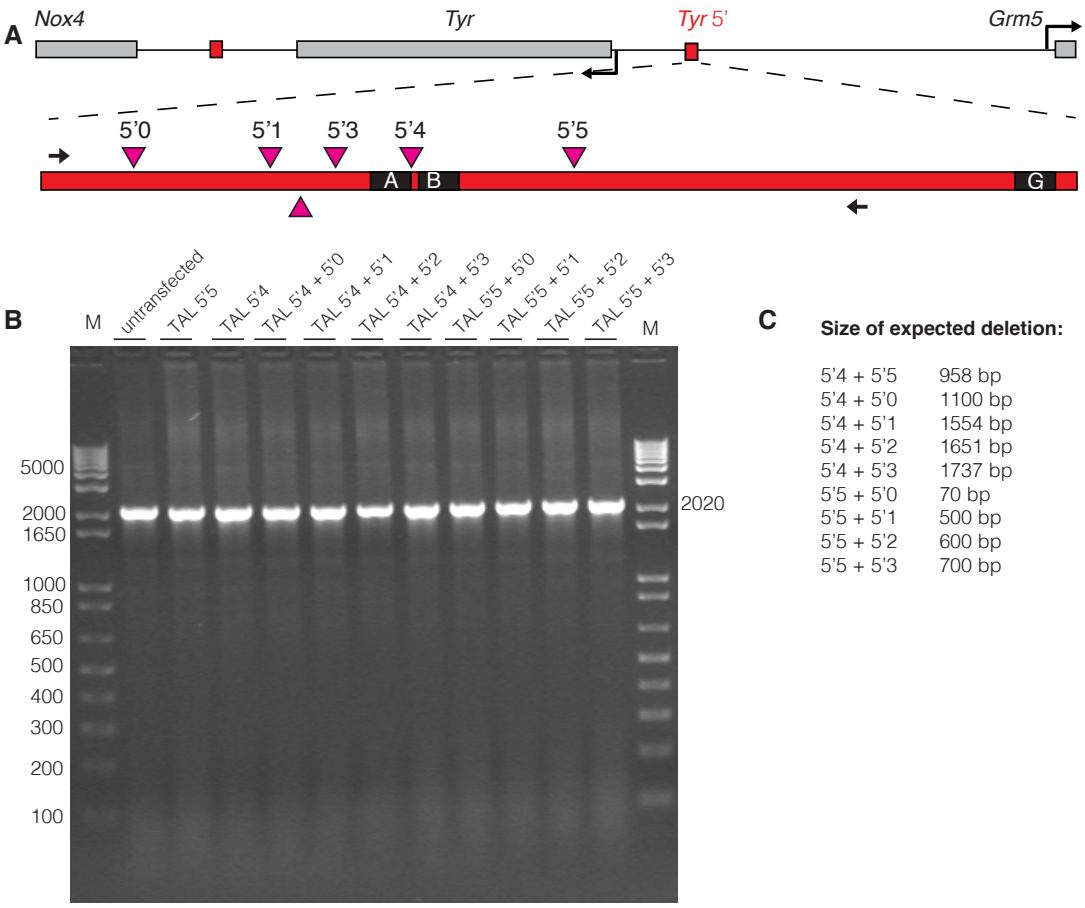


Figure r17 ***In vitro*** **cleavage assay of the Tyr TALENs**. **A** Position of the TALEN target sites across the *Tyr* 3' sequence. The core sequence is indicated as black box. TALENs binding sites are indicated as red triangles **B** Protein extracts of cells transfected with each TALENs pair were incubated with a PCR product containing the corresponding target sequence. Cleavage products are analysed by agarose gel electrophoresis. Each TALE array was cloned and tested in the homodimeri *Fok*I vector and in its heterodimeric derivatives, the mutants ELD and KKR.

TALENs 3'3, 3'5 and 3'6 showed the highest cleavage activity. On the contrary, TALEN 3'1 showed no detectable cleavage. TALEN *Tyr* 3'2 displayed intermediate cleavage activity. Also in this case, obligate-heterodimeric *Fok*I vectors performed better and were chosen for the next experiments (**Figure r17**).

After the validation of individual TALENs using the *in vitro* cleavage assay, we next tested whether we could delete the intervening DNA by co-transfecting two TALEN pairs targeting distal sites flanking each of the two *Tyr* boundaries. To this end, we transfected different combinations of TALEN pairs to mouse Neuro2A cells (Olmsted et al. 1970) that we can transfect with good efficiency.



r2 Figure r18 Transfection of TALEN sets in mouse N2A cells. **A** Position of the different TALEN pairs combinations transfected in Neuro2A cells. TALENs binding sites are indicated as red triangles **B** PCR analysis indicating that no detectable chromosomal deletions are produced in mouse Neuro2A cells. **C** Size of the expected bands in case that deletions are produced using the TALENs combination indicated.

Genomic DNA was collected 72 hours post-transfection and analysed by PCR. Surprisingly, none of the tested TALEN pair combinations resulted in a detectable chromosomal deletion (**Figure r18**). We speculated that the transfection efficiency we obtained in Neuro2A cells could be not as high as necessary. Either TALENs were not expressed at a sufficient level to trigger two simultaneous DSBs, or the number of cells in which the deletion occurred was below the PCR detection limit. We therefore decided to use another cell line, human HEK 293 cells, that we can transfect with very high efficiency. Being HEK 293 a human-derived cell line, we needed to co-transfect a plasmid carrying the full mouse *Tyr* boundaries sequences. For the *Tyr* 5' deletion, we used the pHS6.1 plasmid (Gimenez et al. 2004), containing the complete sequence of the *Tyr* 5' boundary, including target sites of all designed TALENs (**Figure r19 A**) along with *Tyr*

promoter and a *Tyr* minigene. With this approach, we could observe the expected deletion in six out of the nine TALEN combinations tested (**Figure r19 B**).

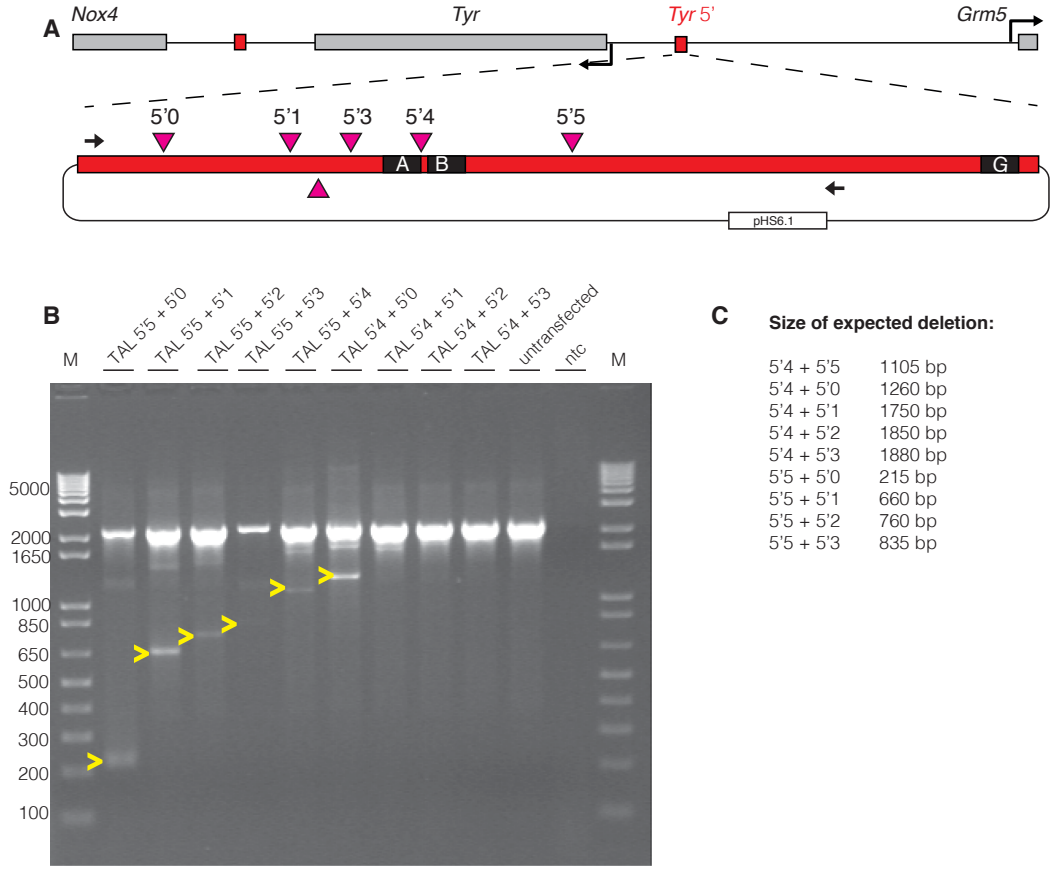


Figure r19 Expected deletions with episomal template (*Tyr* 5') **A** Position of the different TALEN pairs combinations transfected in human HEK 293 cells carrying the pHS6.1 plasmid, containing the mouse *Tyr* LCR sequence. TALENs binding sites are indicated as red triangles **B** PCR analysis indicating that some of the expected deletions are produced, indicated by yellow marks. **C** Size of the expected bands.

Similarly, we used HEK 293 cells to test *in vitro* which TALENs combinations would be the best to delete the *Tyr* 3' boundary. We took advantage of the BAC PR24-276I14 (**Figure r20 A**), containing the *Tyr* locus previously used for the 3C assay. BAC DNA along with different combinations of TALEN pairs were co-transfected in HEK 293 cells. Total DNA was analysed by PCR 72 hours post-transfection. In this experiment, four out of the nine tested combinations successfully produced the expected deletion (**Figure r19 B**). Notably, the TALEN 3'1, that was unable to cleave its target in the in vitro cleavage assay, also failed when combined with other TALENs, confirming the previous results.

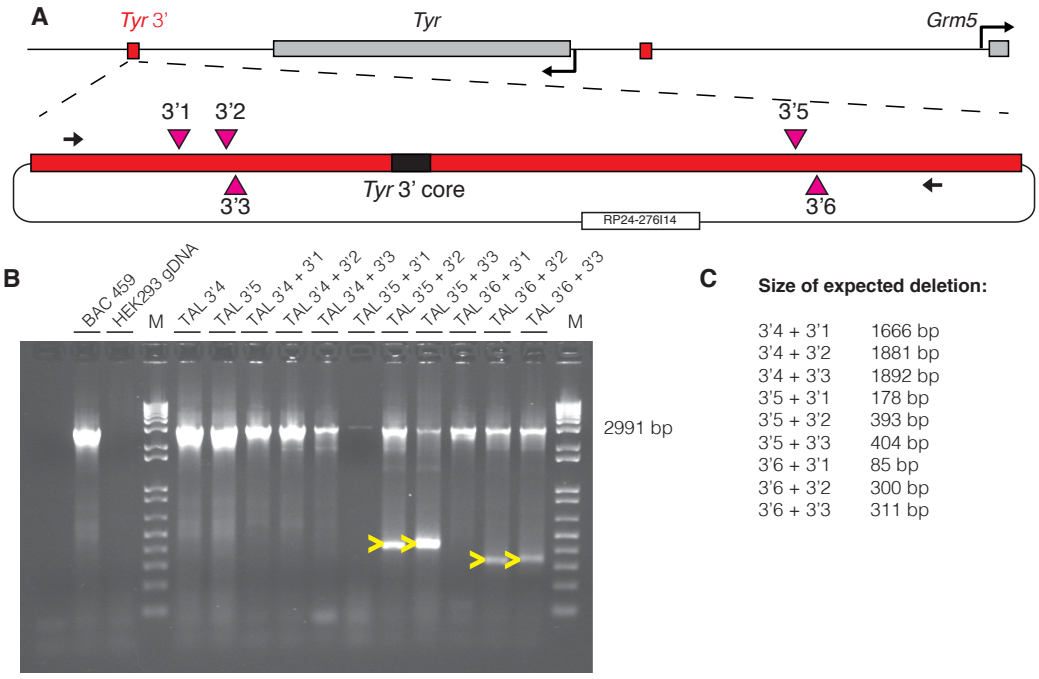


Figure r20 Expected deletions with episomal template (Tyr 3') **A** Position of the different TALEN pairs combinations transfected in human HEK 293 cells carrying the RP24-276I14 BAC, containing the mouse *Tyr* 3' sequence. TALENs binding sites are indicated as red triangles **B** PCR analysis indicating that some of the expected deletions are obtained, indicated by yellow marks. **C** Size of the expected bands.

4.3.3 Deleting *Tyr* 5' boundary *in vivo* with TALENs

As a result of the experiments in cultured cells, we selected two TALEN pairs to be used in combination for the chromosomal deletion of the *Tyr* 5' boundary in mouse. We focused the TALEN set 5'0 + 5'5 that are, among the TALEN combinations successfully tested in cultured cells, the ones that enable the largest deletion (1230 bp) (**Figure r21**). The same vectors used for the transfection in cultured cells were used to produce TALENs mRNA (section 3.5) TALEN mRNA mix was delivered by cytoplasmic injection (Hermann et al. 2014) in mouse fertilized oocytes (C57BL/6J). These microinjection experiments were carried out in the laboratory of Dr. Pawel Pelczar, Institute of Laboratory Animal Science, University of Zürich, Switzerland.

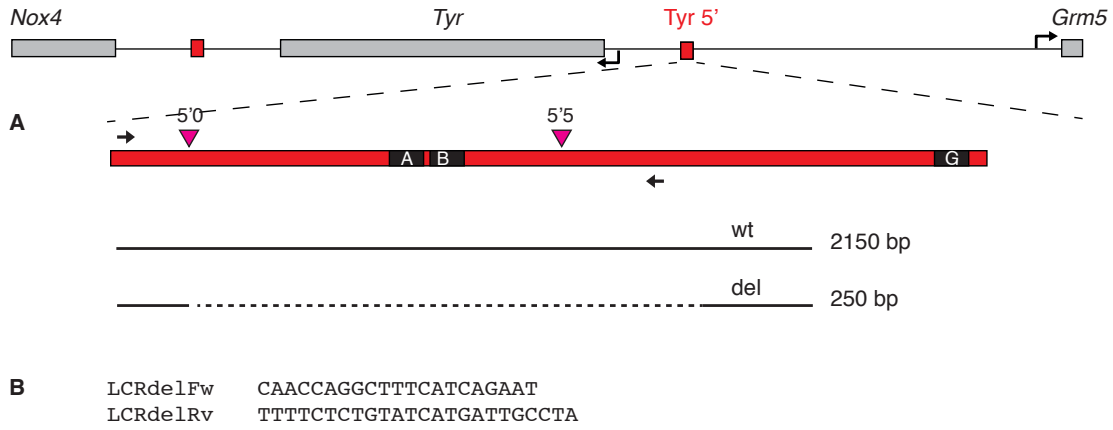


Figure r21 Selected TALEN pairs for *Tyr* 5' deletion *in vivo*. **A** The chromosomal position of each TALEN pair is indicated. A, B and G boxes are indicated as black boxes. TALENs binding sites are indicated as red triangles. **B** Genotyping strategy and sequence of the oligonucleotides used.

Table r2 illustrates the result of the first microinjection session for the Tyr 5' deletion, performed using a 40 ng/μl RNA solution containing 10 ng/μl of each TALEN monomer.

Table r2 Results of TALENs 5'0+5'5 microinjection.

TALEN concentration	Injected	Transferred	Fosters	Pups (Total - stillbirths)
4 x 10 ng/μl	150	97 (64%)	5	39-4 = 35

Out of the 35 live pups obtained, we could not find any positive mice (**Figure r22**).

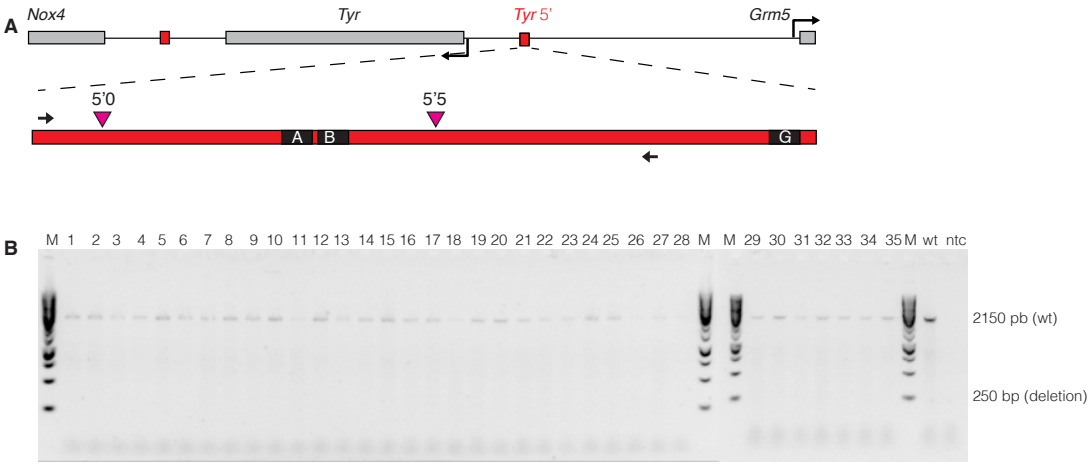


Figure r22 Result of Tyr 5' TALENs microinjection **A** The expected deletion induced by the TALENs 5'0 and 5'5 will reduce the size of the amplified fragment to 215 bp. **B** None of the analyzed mice showed the smaller 215 bp band, nor any other band.

We next asked if at least one of the two nucleases reached its target site and induced any detectable mutations through the NHEJ repair pathway. The TAL 5'0

spacer sequence contains a *Mwo*I restriction site. If TAL 5'0 induced a DSB on its target site, the resulting *indels* would disrupt the *Mwo*I restriction site. We therefore PCR amplified the TAL 5'0 region and digested the PCR with *Mwo*I. As illustrated in **Figure r23**, of the 18 mice analysed, none showed the loss of the *Mwo*I site, indicating no TALEN activity.

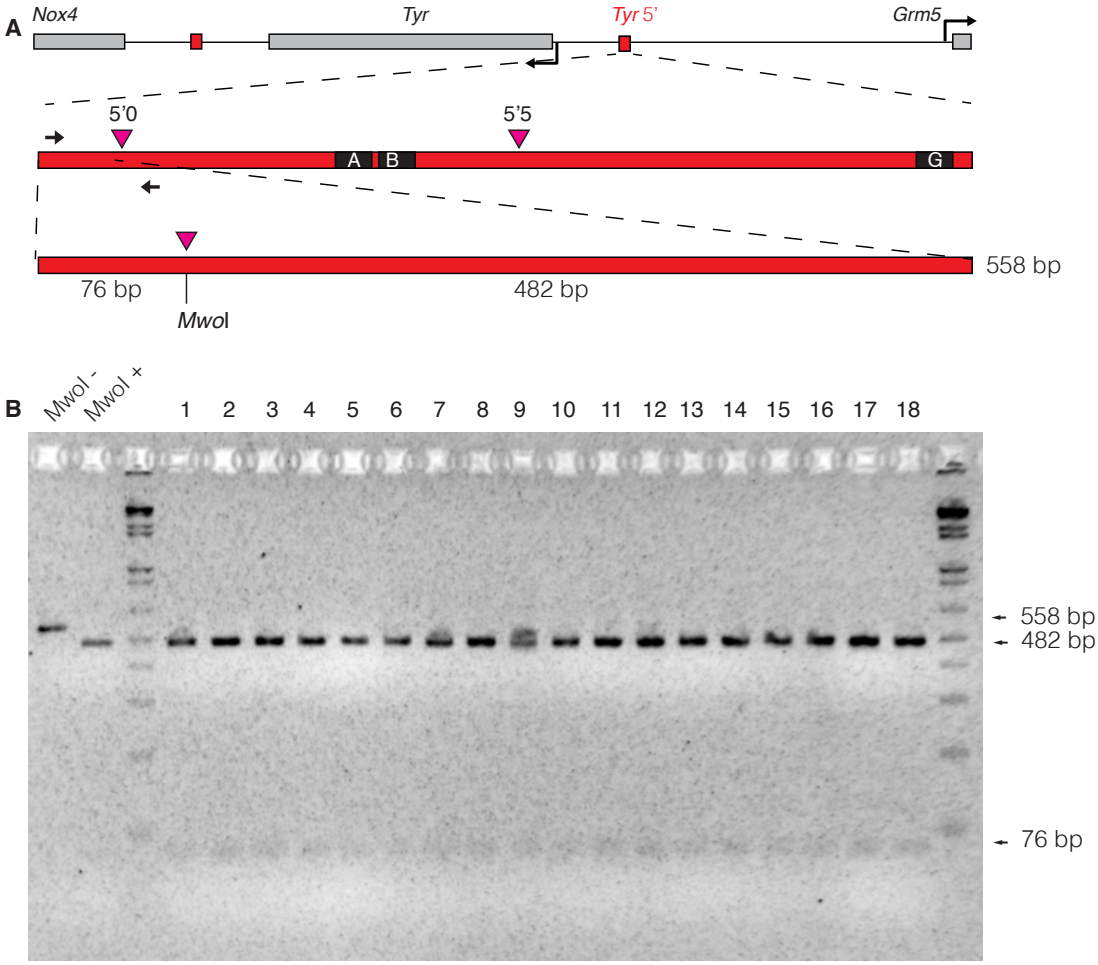


Figure r23. Activity of the TALEN Tyr 5'0 in vivo. **A** *Indels* produced by the TALEN 5'0 disrupts a *Mwo*I restriction site located on the TALEN spacer sequence, resulting in the loss of the restriction site. **B** All PCRs from selected individuals of the TALENs microinjection can be digested by *Mwo*I, suggesting that TALEN 5'0 is not active *in vivo* at the tested concentration. A wild-type control is included in lane 1 (undigested) and 2 (*Mwo*I digested).

The injection of 10 ng/μl of each TALEN mRNA did not produce any mutation in the resulting offspring. We considered increasing the RNA dose and therefore injected 50 ng/μl of each TALEN RNA species, 200 ng/μl of total RNA. Results of this second microinjection session are summarized in **Table r3**.

Table r3 Results of TALENs 5'0+5'5 microinjection/2

TALEN concentration	Injected	Transferred	Fosters	Pups (Total - stillbirths)
4 x 50 ng/µl	205	125 (60%)	6	35

Also from the 50 ng/µl microinjection series we could not obtain any founder carrying the desired deletion, as no band corresponding to a deleted allele could be observed by PCR (**Figure r24**).

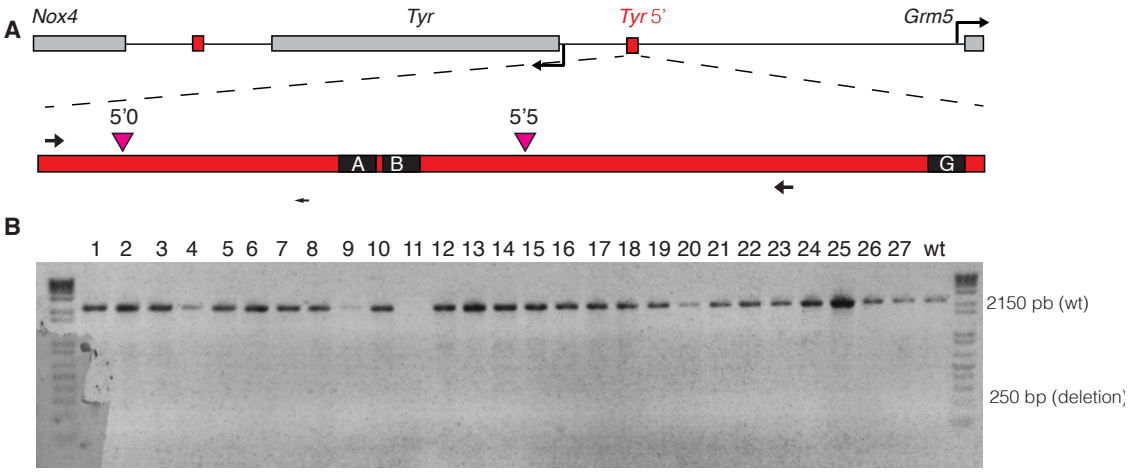


Figure r24 Result of *Tyr* 5' TALENs second microinjection. **A** The expected deletion induced by the TALENs 5'0 and 5'5 will reduce the size of the amplified fragment to 215 bp. **B** None of the analyzed mice showed the smaller 215 bp band, nor any other band.

Similarly to what had been done for the 10 ng/ml injection series, we analysed the activity of the TAL 5'0. In the eight mice analysed, we could not detect any trace of TALEN activity (Figure r25). In conclusion, we could not obtain the *Tyr* 5' deletion using neither of the two different concentrations of TALENs mRNA that we used. In addition, analyses of individual TALEN target sites showed no TALEN activity in all mice that we analysed.

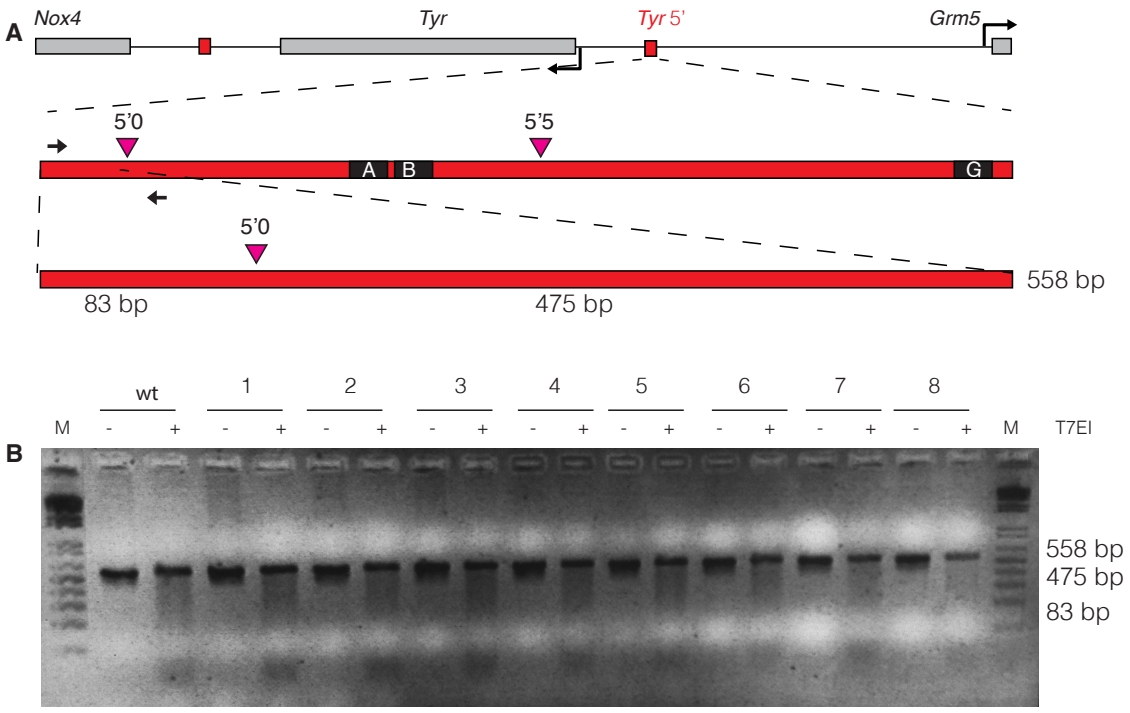


Figure r25 T7 Endonuclease I assay for the TALEN *Tyr* 5'0. **A** Bands of the approximate size of 83 and 475 bp are expected in case of activity of the TALEN 5'0. **B** Digestion of PCR using T7 Endonuclease I did not produced bands, indicating that TALEN 5'0 didn't induced DSB at the targeted site.

4.3.4 Deleting *Tyr* 3' boundary *in vivo* with TALENs

For the *Tyr* 3' deletion *in vivo* we focused on the TALENs combination 3'3 + 3'5 (**Figure r26**). This combination was successfully used to generate a 2570 bp deletion in cultured cells (**Figure r20**).

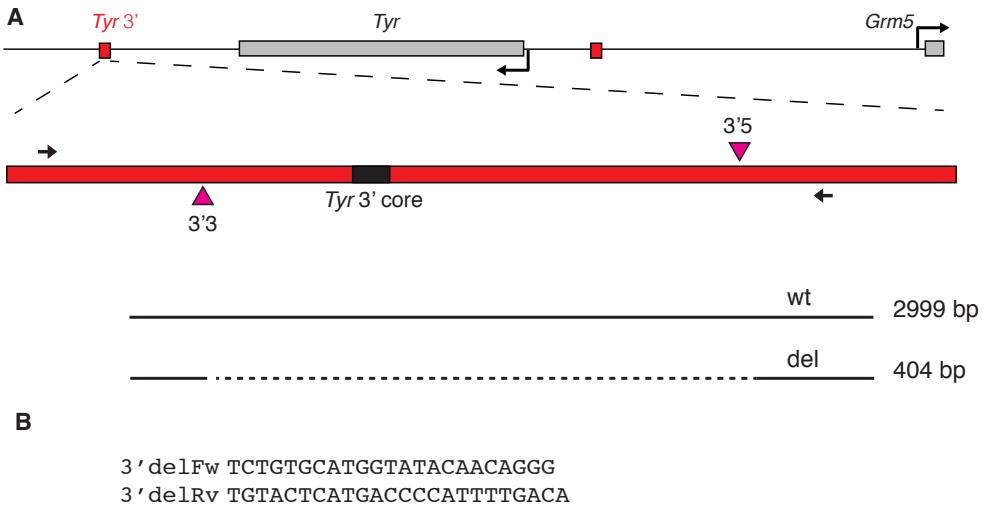


Figure r26: Selected TALEN set for Tyr 3' inactivation in vivo. **A** The chromosomal position of each TALEN pair is indicated. The Tyr 3' core element, composed of a SINEB2 element and a CTCF-binding site, is indicated as a black box. TALENs binding sites are indicated as red triangles. **B** Genotyping strategy and sequence of the oligonucleotides used.

As done for the *Tyr* 5' series, we performed two series of cytoplasmatic microinjection: using a 40 ng/μl RNA solution containing 10 ng/μl of each RNA species (**table r4**) and using a 200 ng/μl RNA solution containing 50 ng/μl of each RNA species (**table r5**).

Table r4 Results of TALENs 3'3 + 3'5 microinjection

TALEN concentration	Injected	Transferred	Fosters	Pups (Total - stillbirths)
4 x 10 ng/μl	135	80 (59%)	4	39-7 = 32

Out of the 32 pups generated in this microinjection session, none carried the desired large chromosomal deletion. A second microinjection session was performed with a 5-fold more concentrated RNA solution. Results are summarized in the following table.

Table r5 Results of TALENs 3'3 + 3'5 microinjection using 50ng/μl of each TALEN mRNA

TALEN concentration	Injected	Transferred	Fosters	Pups (Total - stillbirths)
4 x 50 ng/μl	117	70 (59%)	4	29-5 = 24

No animals carrying the desired (and any other) deletions were obtained from this second microinjection session.

4.3.5 CRISPRs

When we began microinjecting the first combination of TALEN mRNA, the first proof-of-principle experiments of CRISPR/Cas9 mediated genome editing in mice were published (Shen et al. 2013; Wang et al. 2013). We hypothesized that the *Tyr* locus could be efficiently targeted using this new, apparently more powerful tool. According to the first published works (Jinek et al., 2013; Mali et al. 2013), CRISPR binding sites (5'-$N_{(20)}$-NGG) were picked manually on repeat-masked regions flanking the *Tyr* 3' and 5' insulators. As done with TALENs, we designed nucleases targeting the A and B boxes, but not the G box. Targeting the G box would require the production of a large deletion including all the LINE1 element. A custom-made bowtie script (see section 3.6.1) was used to identify potential off-targets sites and select the most specific CRISPR sites.

In **table r6** are reported the results of the *bowtie* analysis indicating the best CRISPR binding sites sequences (**Figure r27**):

Table r6 Selected CRISPR target sites for *Tyr* 3' and 5' deletion. The 20-mer target site is indicated for each region to target. The PAM motif, that is necessary to trigger the CRISPR/Cas9-mediated DNA cleavage is shown in green. Genomic coordinates of the selected CRISPR target sites are given, based on assembly NCBI37/mm9 (July 2007).

ID	target 5'-$N_{(20)}$	PAM	chromosomal position
CRISPR 3'3:	AGCTCAGTAGAGTACTAGGT	AGG	chr7: + 94559177
CRISPR 3'5:	ACAAACGCTAATTGGTAAAA	TGG	chr7: + 94562048
CRISPR 5'0:	GTGTGACAGTGCAAGATAAC	AGG	chr7: + 94656546
CRISPR 5'5:	GTGATAAAACTAGGCAATTT	TGG	chr7: - 94657775

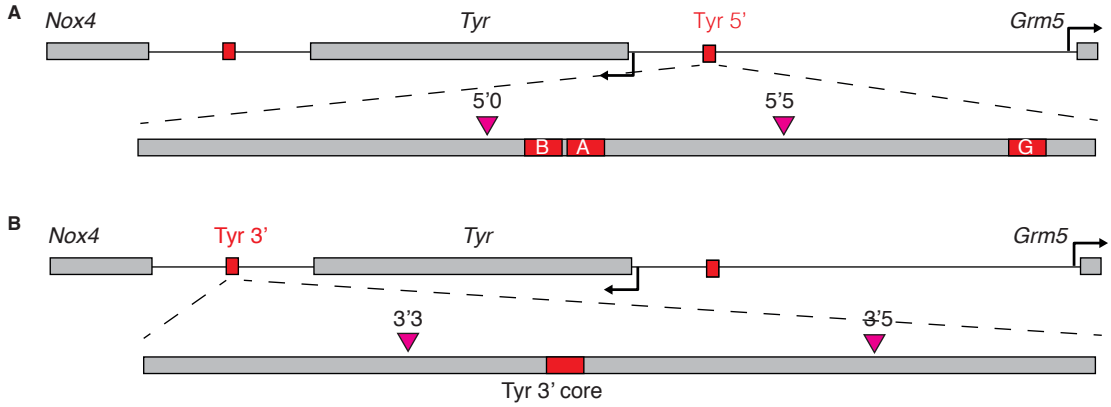


Figure r27 Selected CRISPR sets for *Tyr* 5' and 3' inactivation. **A** Chromosomal position of CRISPR target sites to be used for the deletion of *Tyr* 5' boundary. Contemporary cleavage of CRISPR 5'0 and 5'5 will result in a 1.2 kb deletion. **B** Chromosomal position of CRISPR target sites to be used for the deletion of *Tyr* 3' boundary. Contemporary activity of CRISPR 3'3 and 3'5 will produce a 2.9 kb deletion.

These CRISPR binding sites were further analysed using the CRISPR design tool developed by the Feng Zhang laboratory (crispr.mit.edu). This tool assigns a score to each predicted off-target, ranging from 0.1 to 100, taking in consideration the number and the position of mismatches with the desired target sequence. The higher the score, the higher the probability that the off-target might be targeted by the CRISPR in study. This second search confirmed the good quality of the selected CRISPR sites (the complete list of predicted off-target sites can be found in the **Appendix 9.2**).

4.3.6 Deleting *Tyr* boundaries in vitro with CRISPRs

Four sgRNA (small guide RNA) expression vectors needed to transcribe the small RNA molecule homologous to the targeted sites (described in section 3.1.5)

were produced for each of the selected CRISPR targets, as described in section 3.1.11, by Golden Gate Cloning (3.1.9).

Mouse Neuro2A cells were transfected with each sgRNA pair, along with a Cas9 expression vector, as described in section 3.6.4. The presence of the expected deletion was tested by PCR 72 hours post-transfection using primers mapping outside the targeted sequences. Indeed, the co-transfection of Cas9 and two different sgRNA, mapping on distal sequences on the same chromosome, produced the deletion of the intervening DNA, (**Figure r28** and **r29**). Hence, using CRISPRs, but not TALENs, we could successfully delete the *Tyr* 5' and the *Tyr* 3' boundaries in the genome of mouse Neuro2A cells.

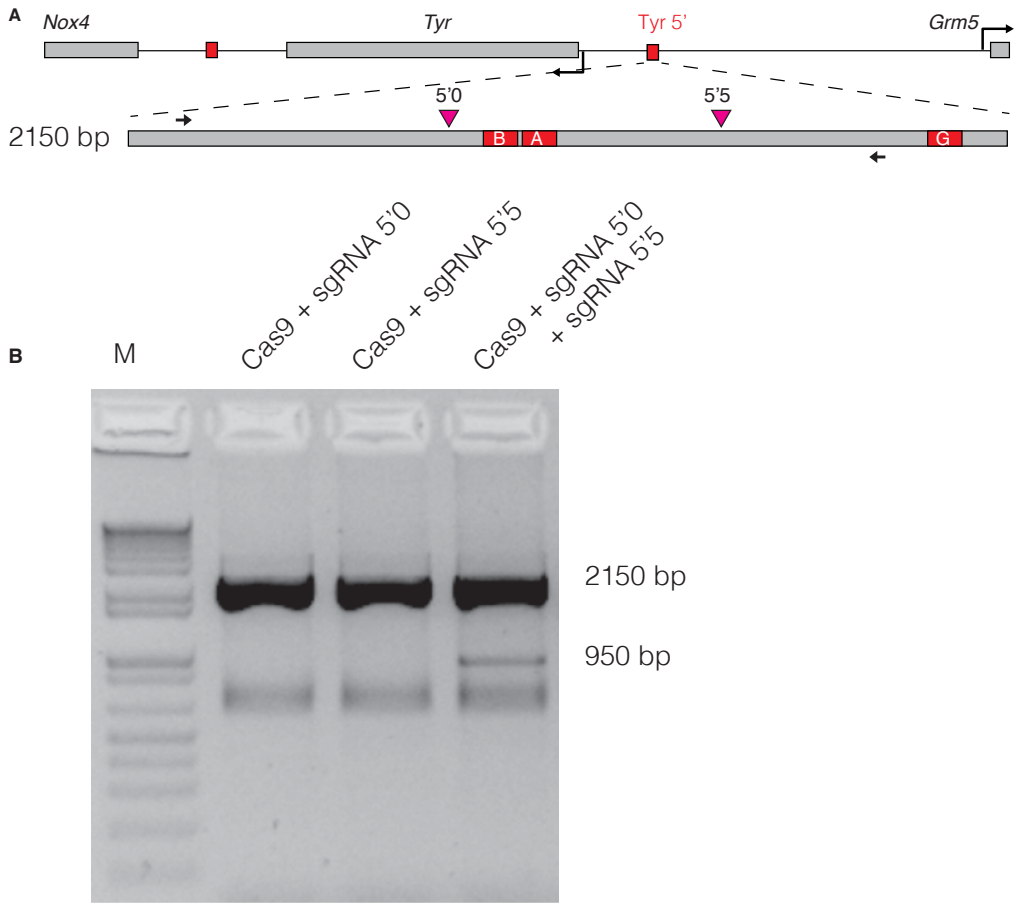


Figure r28: Targeted deletion in mouse Neuro2A cells with CRISPR/Cas9 (*Tyr* 5'). **A** Position of the CRISPR target sites (indicated as red triangles) on the mouse *Tyr* 5' sequence. Simultaneous cuts at both target sequence results ina 1.2 kb large deletion. Only when both sgRNA expression vectors are co-transfected alogn with a Cas9 expression vector, the expected deletion occurs.

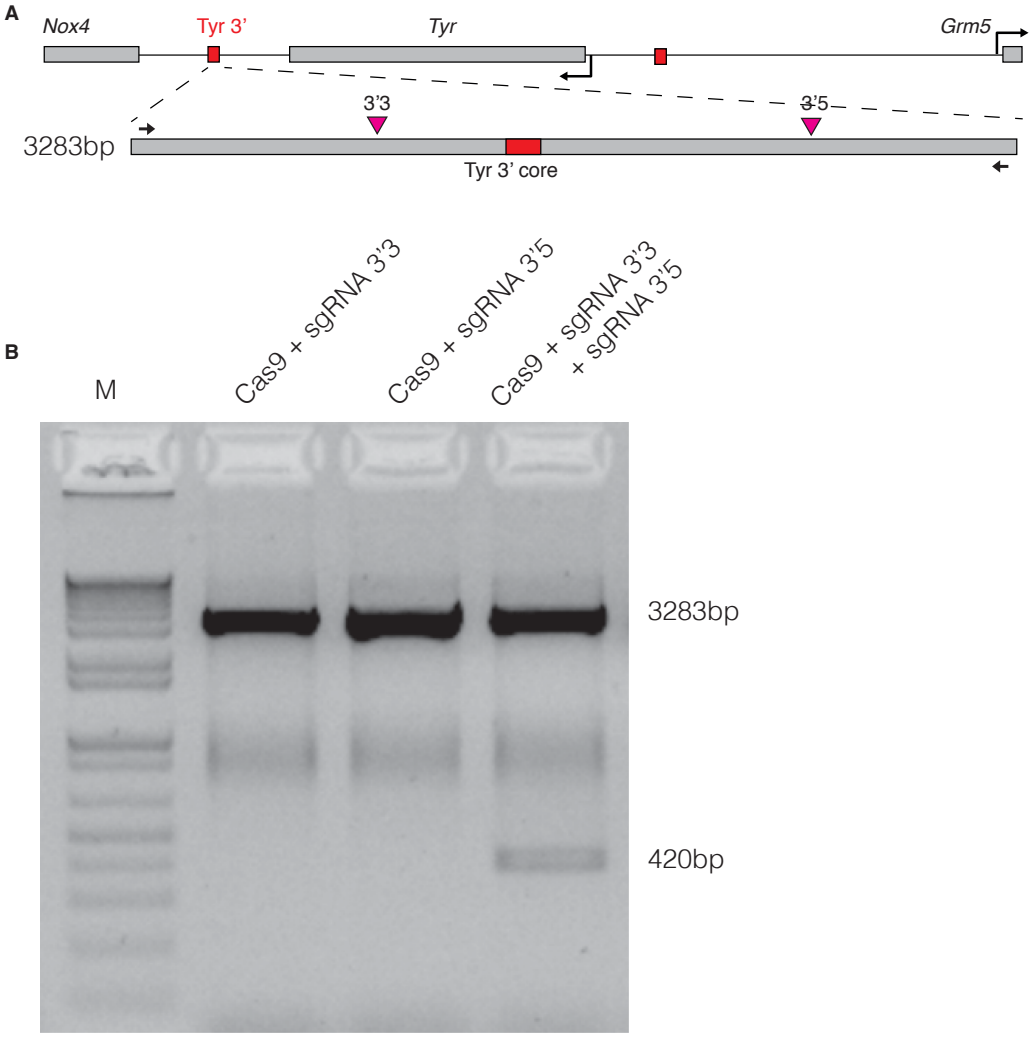


Figure r29: Targeted deletion in mouse Neuro2A with CRISPR/Cas9 (Tyr 3'). **A** Simoultaneous cuts would induce the deletion of the 2.8kb intervening sequence. **B** A 420 bp band corresponding to the targeted deletion of the mouse *Tyr* 3' insulator is observed only when both sgRNA expression vectors and Cas9 expression vector are transfected.

4.3.7 Deleting *Tyr* 5' boundary *in vivo* with CRISPRs

After *in vitro* validation, we adapted the CRISPR expression vectors for the delivery into mouse fertilized eggs. By PCR, we added the T7 RNA polymerase promoter sequence 5' to the Cas9 ORF and produced 5' capped, 3' poly A-tailed mRNA (see section 3.6.4 for a detailed protocol). Similarly, by PCR we obtained a 120 bp DNA fragment composed of 100 bp of target-specific sgRNA preceded by the T7 promoter sequence. sgRNA PCR products were used for RNA *in vitro* transcription without capping and polyA tailing. The first microinjections were performed using 20 ng/μl Cas9 mRNA and 10 ng/μl of each of the two sgRNA (sgRNA 5'0 and sgRNA 5'5). **Table r7** illustrates the results of the first three

microinjection sessions performed in the laboratory of Pawel Pelczar, University of Zurich, Switzerland performed on B6D2F1 genetic background.

Table r7 Results of CRISPR 5'0 + 5'5 microinjection

RNA concentration	Injected	Transferred	Fosters	Pups (Total - stillbirths)
20 ng/µl + 2 x 10 ng/µl	120	62 (51%)	3	20
20 ng/µl + 2 x 10 ng/µl	120	60 (50%)	3	13
20 ng/µl + 2 x 10 ng/µl	135	84 (70%)	3	10
TOTAL	**510**	**290 (56%)**	**9**	**43**

A total of 43 pups were obtained and genotyped by PCR. A 2150 bp band is expected for wild-type animals, whereas a 950 bp is expected if the deletion is produced correctly. None of the analysed mice showed the 950 bp band corresponding to the expected deletion, although animal #40 showed a smaller PCR product, probably indicating a partial deletion (**Figure r30**)

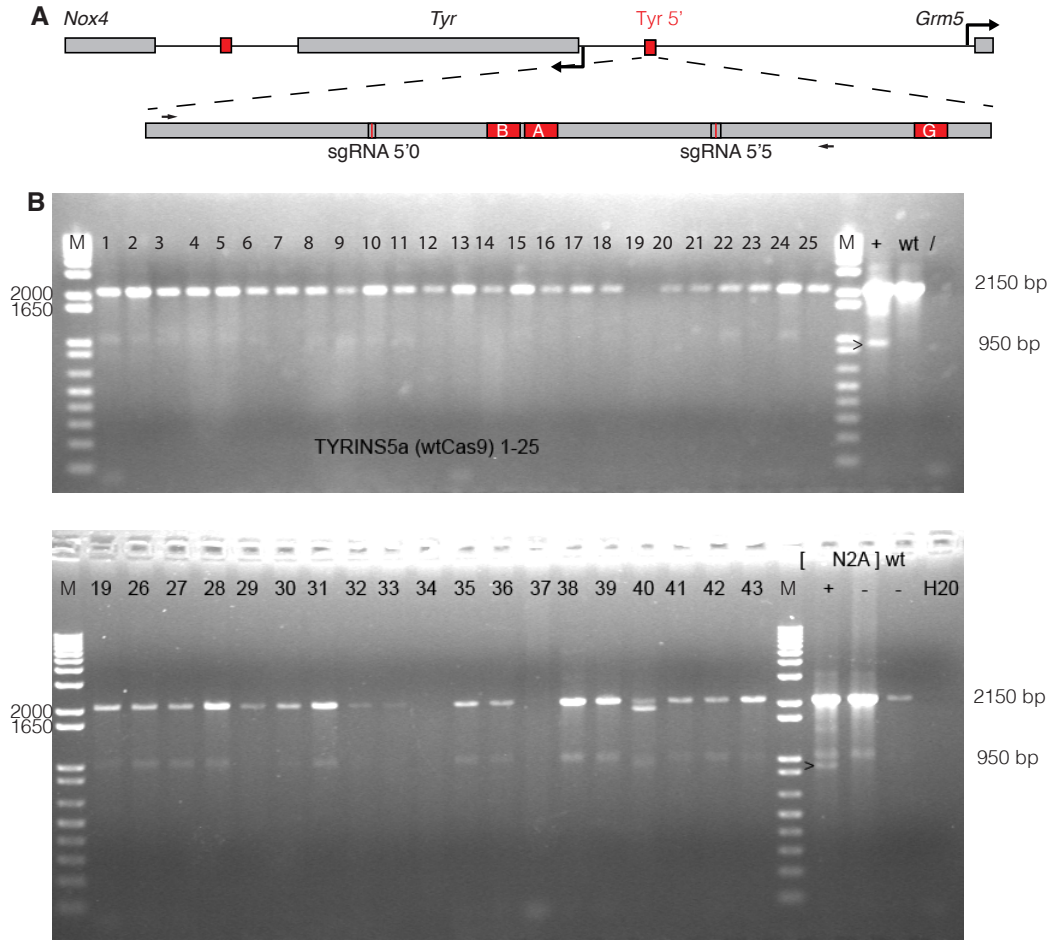


Figure r30 Genotyping of animal obtained from CRISPR 5'0 + 5'5 microinjection. **A** A 2150 bp band is expected from the full *Tyr* 5' amplification. Upon CRISPR/Cas9-mediated deletion, a smaller 950 bp band is produced. **B** Genotyping by PCR. N2a cells transfected with CRISPR/Cas9 plasmids are used as positive control. Untransfected N2a cells and DNA from a wild-type mouse were used as negative, wild-type control. None of the analysed mice showed the expected deletion pattern. Mouse #40 showed a smaller band, probably indicating a small, partial deletion derived by cellular exonuclease activity following the CRISPR-induced DSB.

The partial deletion was confirmed by cloning and Sanger sequencing (**Figure r31**): 202 bp of genomic DNA were deleted in mouse #40.

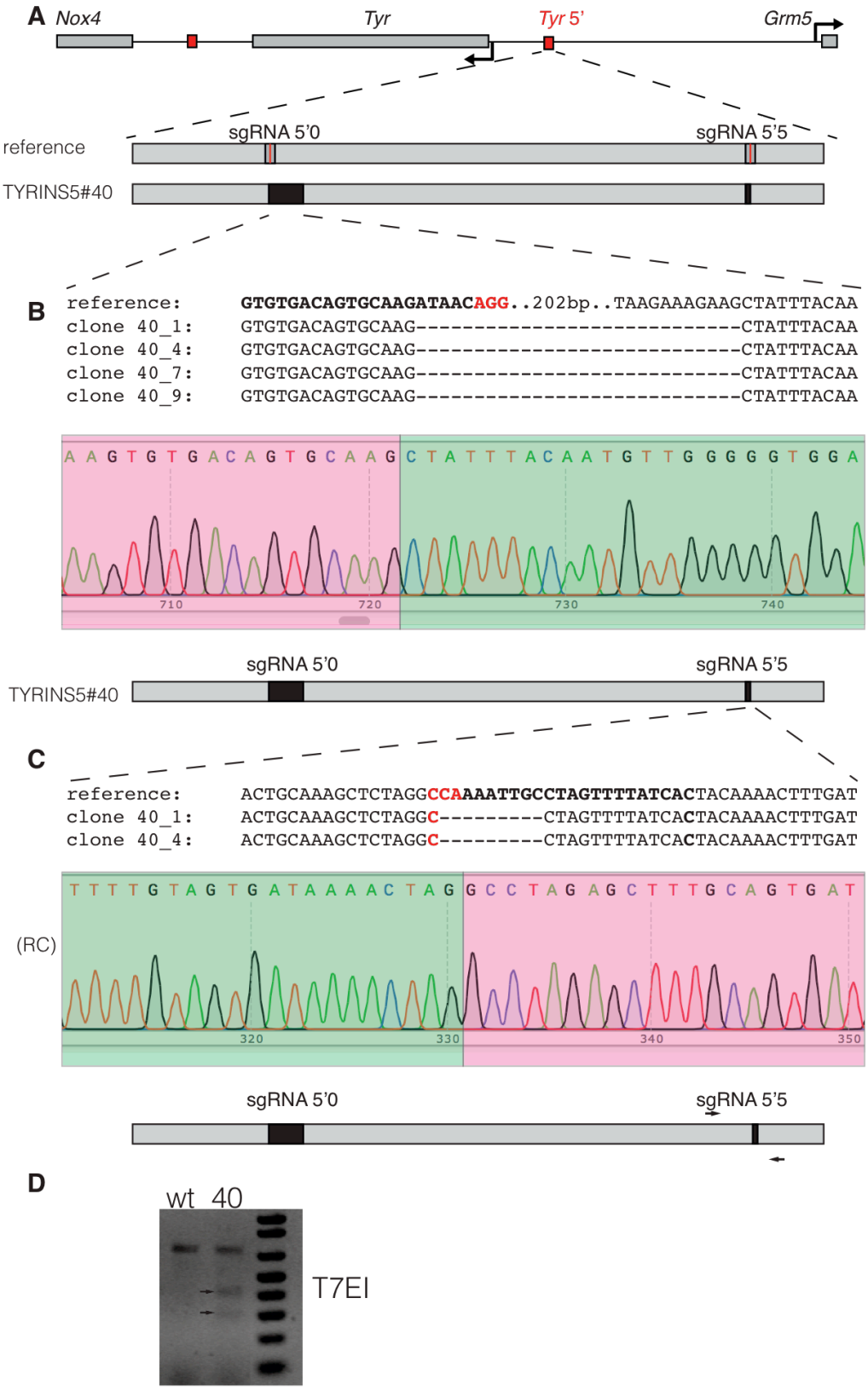


Figure r31 Characterization of animal TYRINS5 #40. **A** sgRNA position with respect of the *Tyr* locus and the *Tyr* 5' element. **B** A 202 bp deletion was confirmed by Sanger Sequencing, starting from 4 bp upstream the PAM of the CRISPR 5'0 binding site. The finding was confirmed in 4 independent clones. **C** Surprisingly, footprint of CRISPR activity was found, on the same allele, in correspondence of the CRISPR

5'5 binding site. Sequencing of two independent clones showed a 9 bp deletion. **D** The indels at the CRISPR 5'5 site were also confirmed by T7 Endonuclease I assay.

The observation of the partial deletion in animal #40 represents the first proof of activity of targeted nucleases *in vivo*, in this work. In animal TYRINS5#40, the *Tyr* 5' boundary sequence was targeted by both CRISPRs on the same allele, but unfortunately, not simultaneously.

We proceeded with a second series of microinjections, using a higher concentration of RNA. The Cas9 mRNA was raised up to 25 ng/μl, and the sgRNA was increased from 10 ng/μl to 45 ng/μl.

RNA concentration	Injected	Transferred	Fosters	Pups (Total - stillbirths)
25 ng/μl + 2 x 45 ng/μl	239	104 (43%)	4	12
25 ng/μl + 2 x 45 ng/μl	162	88 (54%)	4	28
TOTAL	**401**	**192 (47%)**	**8**	**40**

Table r8 Results of CRISPR 5'0 + 5'5 microinjection/2

We obtained a total of 40 animals. By PCR, we concluded that none of them carried the desired deletion. Analysis of individual nuclease activity, by T7 Endonuclease I assay at the CRISPR 5'0 and 5'5 site, showed no nuclease activity (**Figure r32** and **r33**). Considering that we could indeed detect nuclease activity for both CRISPRs 5'0 and 5'5 in the previous injection series (**Figure r31**), we concluded that a technical problem could have occurred in this second injection series.

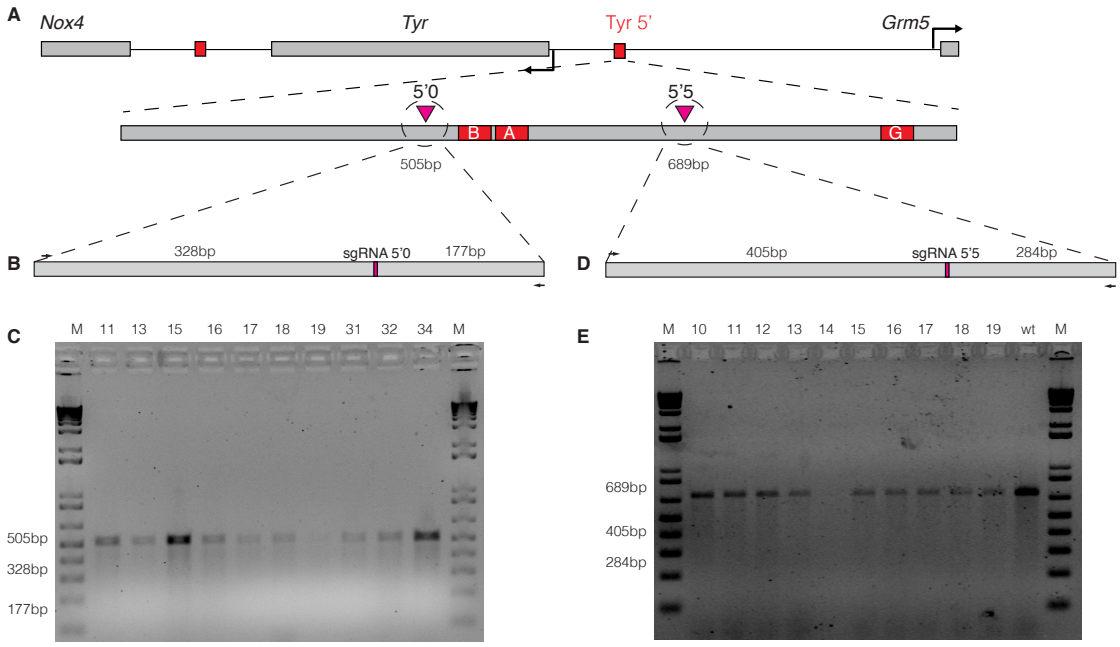


Figure r32 T7 Endonuclease I assay. A Position of the CRISPR target sites with respect of the *Tyr* locus. **B** A 505 bp sequence centred on the sgRNA 5'0 site was PCR-amplified using the oligonucleotides. The region surrounding the CRISPR 5'5 binding site was PCR amplified with the oligonucleotides T7EILCR123Fw and T7EILCR123Rv. In *indels* are present at the sgRNA 5'0 site, the T7 Endonuclease I digestion will produce a 328 bp and 177 bp bands. **C** We could not detect any *indel* in all tested individuals. **D** Similarly, a 689 bp region corresponding to the CRISPR 5'5 site was PCR amplified with the oligonucleotides 67Fw and LCRdelRv. A 405 bp and 284 bp bands are expected when *indels* are present at the expected position. **E** All tested animals are negative also for the T7 Endonuclease I assay for the sgRNA 5'5 site.

According to the results reported in the latest publications (Wang et al. 2013; Yang et al. 2013), we decided to increase the concentration of the RNA mix (50 ng/μl Cas9 mRNA and 90 ng/μl of each sgRNA; 35 ng/μl Cas9 mRNA and 60 ng/μl of each sgRNAs). These microinjection experiments were performed in Madrid at the CNB-CBMSO Transgenic facility in pigmented B6.CBA genetic background

Table r9 Results of CRISPR 5'0 + 5'5 microinjection/3

RNA concentration	Injected	Transferred	Fosters	Pups (Total - stillbirths)
50 ng/μl + 2 x 90 ng/μl	88	32 (36%)	2	5
50 ng/μl + 2 x 90 ng/μl	137	67 (48%)	5	28
35ng/μl + 2 x 60 ng/μl	163	86 (52%)	5	31
TOTAL	388	185 (47%)	12	64

Finally, animals carrying the desired deletion were obtained in the first microinjection session using 50 ng/μl Cas9 mRNA and 90 ng/μl of each sgRNAs (**Figure r33**). Partial deletions were also obtained.

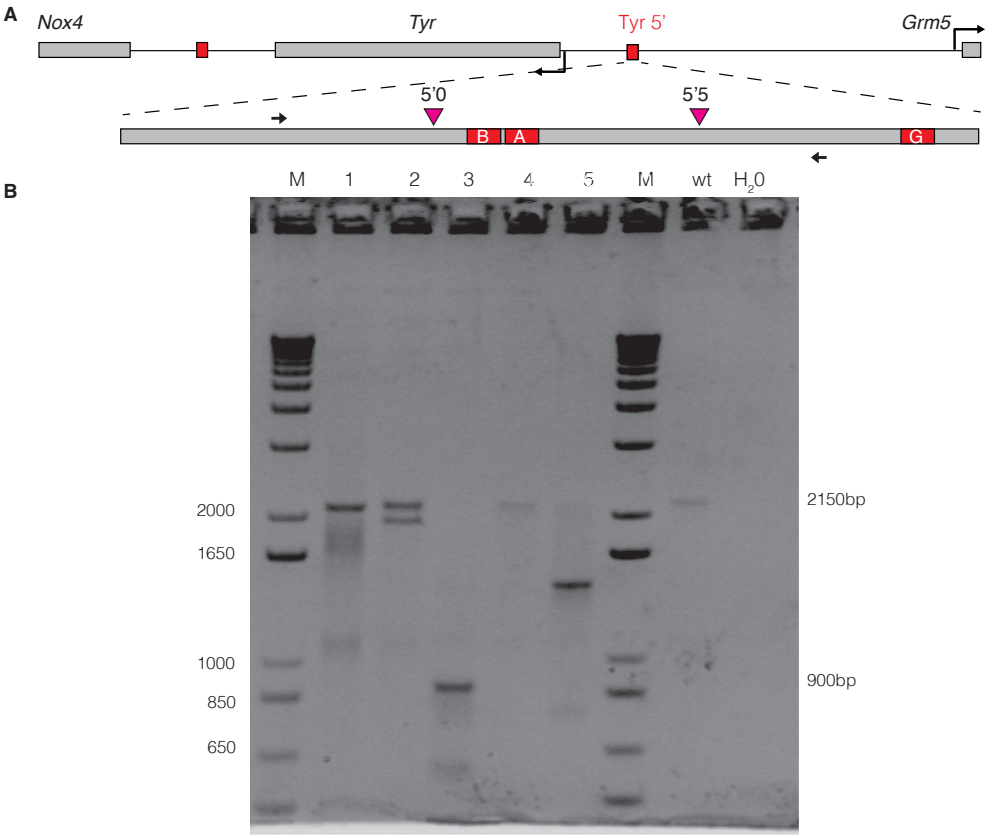


Figure r33 Successful deletion using higher RNA concentration. **A** Position of the CRISPR target sites with respect of the *Tyr* locus. **B** Genotyping of the five individuals obtained from the microinjection session using 50 ng/µl Cas9 along with 90 ng/µl of each sgRNA. Animal 3 shows a band of about 900 bp, compatible with the expected deletion. Animal 5 and 2 show bands suggesting that partial, smaller deletions occurred. Animals 1 and 4 do not show additional bands.

By cloning-sequencing we confirmed the full deletion of individual #3 (**Figure r34**). This animal carries a 1274 bp deletion including the A and B boxes of the *Tyr* LCR. The breakpoint is observed between 81 bp upstream the sgRNA 5'0 site (**Figure R34 A**) and the sgRNA 5'5 site, with an insertion of 76 bp (**Figure R34 B and C**) that does not match with any genomic sequence associated with the mouse *Tyr* locus.

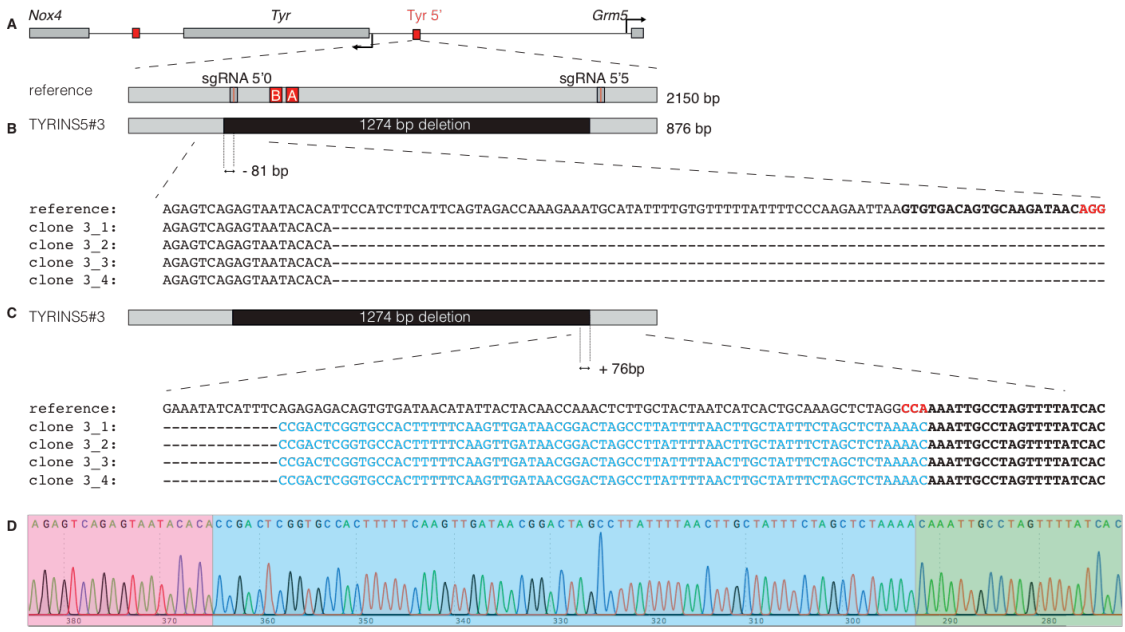


Figure r34 Sequence characterization of animal TYRINS5 #3. **A** Position of the CRISPR target sites with respect of the *Tyr* locus. **B** Sequence alignment of the deletion break-point at the sgRNA 5'0 site. Four independent clones were sequenced and aligned with the reference genome (NCBI37/mm9). The deletion spanned some 1274 bp including 81 bp upstream to the sgRNA 5'0 target site. sgRNA binding site is shown in bold, PAM motif in red. **C** Alignment of the deletion break-point at the sgRNA 5'5 site. The nuclease cut happened exactly downstream the PAM motif (in red) and resulted in a 76 bp insertion (shown in blue). **D** The chromatogram of the region is given. The region 5' upstream the deletion is highlighted in red, the inserted nucleotides are shown in blue. The 3' break-point sequence is shown in red

From the second microinjection session using 50 ng/μl Cas9 mRNA and 90 ng/μl of each sgRNA we obtained 28 live pups. In nine of these individuals deletions were detected at the targeted site (**Figure r35**). Of these, six individuals carry the full 1250 bp deletion of the *Tyr* 5' boundary element (#6; #11; #18; #20; #21 and #30). In the founder #18, in addition to the band corresponding to the full deletion, we could observe a second larger band. This band represents a second, smaller deletion present in this individual. Nevertheless, we could not detect such allele when cloning and sequencing a PCR product obtained from the founder. Indeed, we could sequence it when the allele segregated in the F1 generation. Two individuals carry partial deletions (#8 and #16). Interestingly, one individual carries a deletion larger than expected (#13).

From the third microinjection session (performed using 35 ng/μl Cas9 mRNA and 60 ng/μl of each sgRNA) we obtained 31 animals, of which seven carry deletions in the targeted sequence. Five individuals carry a deletion of the expected size and two animals carry a slightly bigger deletion.

A summary of the results obtained, including schematic and sequencing of all the deletions generated is shown in **Figure r36.**

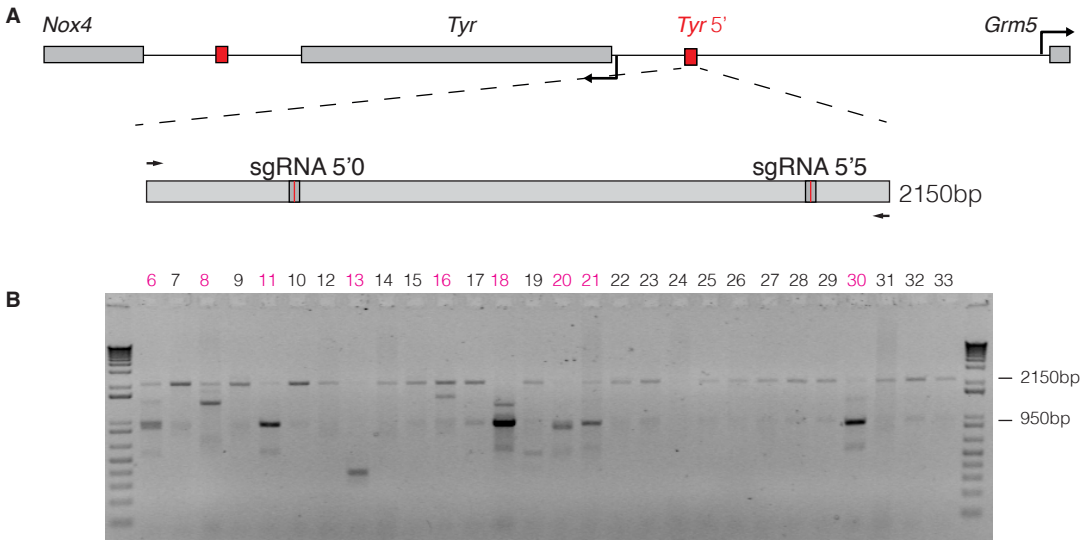
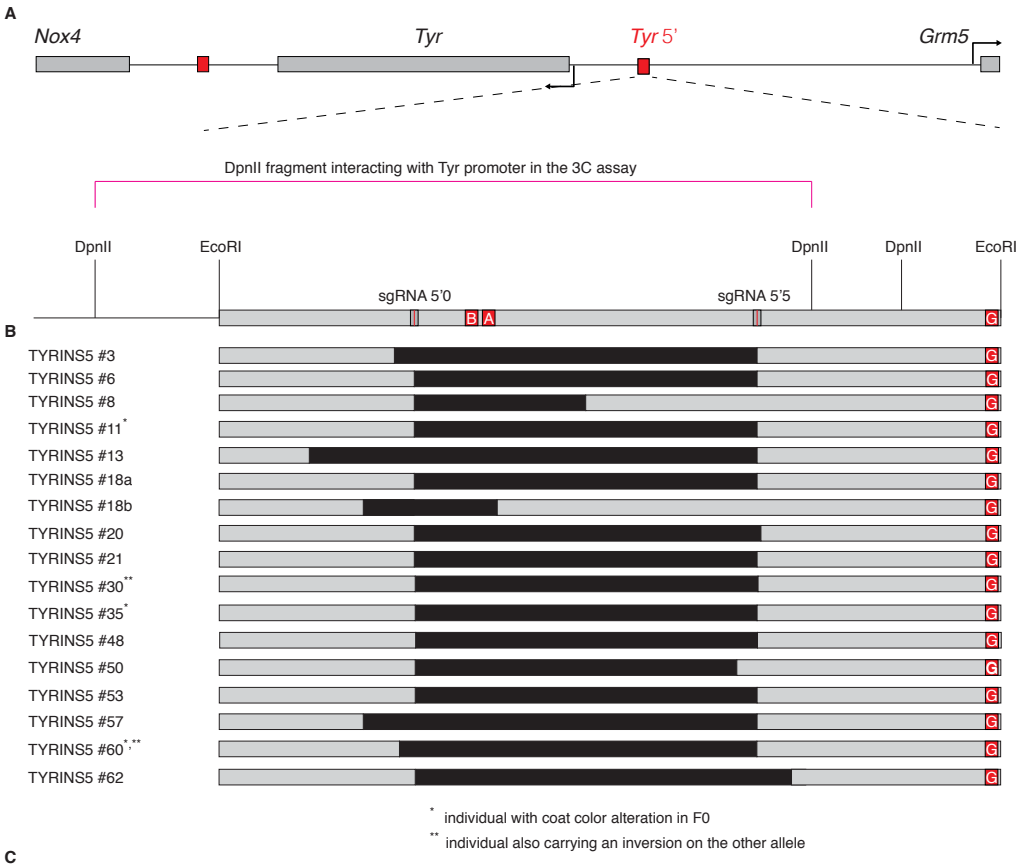


Figure r35 Successful deletion using higher RNA concentration/2. **A** Position of the CRISPR target sites with respect of the *Tyr* locus. **B** Genotyping of the offspring of the second microinjection session using 50ng/µl Cas9 along with 90 ng/µl of each sgRNA. Nine out of 28 individuals carry deletions of various sizes. Six individuals carry a deletion compatible with the expected size. In three individuals, smaller deletion could be observed. Individual #13 show a smaller band suggesting that a larger deletion may have occurred.



C

```
reference:    CAAGAATTAAGTGTGACAGTGCAAGATAACAGGAAAATAA.......//...........CTCTAGGCCAAAATTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#3_1: ----------------------------------------(del 1294bp)----------AAATTGCCTAGTTTTATCACCTATAGTGTT
TYRINS5#3_2: ----------------------------------------(del 1294bp)----------AAATTGCCTAGTTTTATCACCTATAGTGTT
TYRINS5#6_1: CAAGAATTAAGTGT----------------------------(del 1212bp)----------------------------------ACTACAAAACTT
TYRINS5#6_2: CAAGAATTAAGTGTGACAGTGCAAGAAGTTTTTGC---------(del 1151bp)----------AAACACTTGCCTAGTTTTATCACTACAAAA
TYRINS5#8_1: CAAGAATTAAGTGTGACAGTGCAAGGTCGA-------------(del  728bp).....CTCTAGGCCAAAAGTTTATCACTTAATTGCCTAGTTTTAT
TYRINS5#8_2: CAAGAATTAAGTGTGACAGTGCAAGGTCGA-------------(del  728bp).....CTCTAGGCCAAAAG------------TTGCCTAGTTTTAT
TYRINS5#11_1:CAAGAATTAAGTGTGACAGTGCAAGATG---------------(del 1196bp)----------TTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#11_2:CAAGAATTAAGTGTGACAGTGCAAGATG---------------(del 1196bp)----------TTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#13_1:----------------------------------------(del 1783bp)----------CCTAGTTTTATCACTACAAAACTT
TYRINS5#13_2:----------------------------------------(del 1783bp)----------CCTAGTTTTATCACTACAAAACTT
TYRINS5#18a1:CAAGAATTAAGTGTGACAGTGCAAGAT----------------(del 1196bp)----------TTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#18a2:CAAGAATTAAGTGTGACAGTGCAAGAT----------------(del 1196bp)----------TTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#18b1:----------------------------------------(del  784bp).....CTCTAGGCCAAAATTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#20_1:CAAGAATTAAGTGTGACAGTG----------------------(del 1268bp)----------------------------------------
TYRINS5#20_2:CAAGAATTAAGTGTGACAGTG----------------------(del 1268bp)----------------------------------------
TYRINS5#21_1:CAAGAATTAAGTGTGACAGTGC---------------------(del 1205bp)----------CTAGTTTTATCACTACAAAACTT
TYRINS5#21_2:CAAGAATTAAGTGTGACAGTGC---------------------(del 1205bp)----------CTAGTTTTATCACTACAAAACTT
TYRINS5#30_1:CAAGAATTAAGTGTGACAGTGCAAGA-----------------(del 1201bp)----------CTAGTTTTATCACTACAAAACTT
TYRINS5#30_2:CAAGAATTAAGTGTGACAGTCCAAGA-----------------(del 1201bp)----------CTAGTTTTATCACTACAAAACTT
TYRINS5#35_1:CAAGAATTAAGTGTGACAGTGC---------------------(del 1205bp)----------CTAGTTTTATCACTACAAAACTT
TYRINS5#48_1:CAAGAATTAAGTGTGACAGTGCAAGAT----------------(del 1196bp)----------TTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#48_2:CAAGAATTAAGTGTGACAGTGCAAGAT----------------(del 1196bp)----------TTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#50_1:CAAGAATTAAGTGTGACAGTGCAAGATATTTAG----------(del 1097bp).....CTCTAGGCCAAAATTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#50_2:CAAGAATTAAGTGTGACAGTGCAAGATATTTAG----------(del 1097bp).....CTCTAGGCCAAAATTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#53_1:CAAGAATTAAGTGTGACAGTGCAAGAT----------------(del 1196bp)----------TTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#53_2:CAAGAATTAAGTGTGACAGTGCAAGAT----------------(del 1196bp)----------TTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#57_1:----------------------------------------(del 1437bp)----------TTTTATCACTACAAAACTT
TYRINS5#57_2:----------------------------------------(del 1437bp)----------TTTTATCACTACAAAACTT
TYRINS5#60_1:----------------------------------------(del 1297bp)----------ATTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#60_2:----------------------------------------(del 1297bp)----------ATTGCCTAGTTTTATCACTACAAAACTT
TYRINS5#62_1:CAAGAATTAAGTGTGACAGTGCAAG------------------(del 1416bp)----------------------------------------
TYRINS5#62_2:CAAGAATTAAGTGTGACAGTGCAAG------------------(del 1416bp)----------------------------------------
```

Figure r36 Founder individuals of the TYRINS5 line. **A** Position of the CRISPR target sites with respect of the *Tyr* locus. The *Dpn*II fragment interacting with the *Tyr* promoter (in the 3C assay) is highlighted. The position of A, B and G boxes is indicated along with a basic restriction map. **B** Position and dimension of the deletion for each positive individual obtained. The deleted DNA region is indicated as black box. Fourteen animals show a deletion the closely approximate the expected size. One individual carries a smaller, 700 bp deletion centred on the A and B boxes. One animal shows a larger deletion produced by loss of DNA 3' to the sgRNA 5'0 site. **C** Sequence alignment of two independent clones per individual. sgRNA binding sites are shown in bold, PAM motif in red. Inserted bases are shown in blue. Alignments refer to the NCBI37/mm9 mouse genome assembly.

In summary, using the CRISPR/Cas9 system to target the *Tyr* 5', we produced a battery of 14 individual carrying a deletion approximately of the expected size, plus one individual carrying a smaller deletion (which removes the AB boxes plus additional flanking sequences) and another carrying a larger deletion. Another two animals bear smaller deletions, in the range of 100-200 bp, resulting from the nuclease activity of only one of the two nucleases. The most interesting alleles are found in founder #13, that bears the largest deletion, in animals #8, carrying a 0.7 kb deletion centred the AB boxes, and in animal #18. In addition to the full deletion, this founder bears on the other allele a 0.7 kb deletion, inactivating the AB box and other sequences 3' upstream. The comparison of the phenotype obtained in these distinct lines will help to narrow the size of the functional *Tyr* 5' core sequence.

Careful sequencing of the deleted alleles indicated that each line represents an independent event of DNA break and repair, with differences in the range of single bases to tens and hundreds of bases. This reflects the error-prone repair mechanism at the base of this technique, the non-homologous end-joining repair pathway (Barnes et al. 2001). Interestingly, the same break-point sequence is found in animal #18, 48 and 53.

4.3.8 Inverting *Tyr* 5' boundary *in vivo* with CRISPRs

Next, we asked if, in addition to the desired mutation, we have been able to induce also the inversion of the DNA sequences included between the two sgRNA sites, being this a possible resolution of two distal DSBs, as published by Xiao and colleagues (Xiao et al. 2013). By PCR, using a forward oligonucleotide mapping outside the targeted region, in combination with a second forward oligonucleotide mapping inside the targeted region, we could detect specific inversion of the targeted sequence in seven individuals (**Figure r37**).

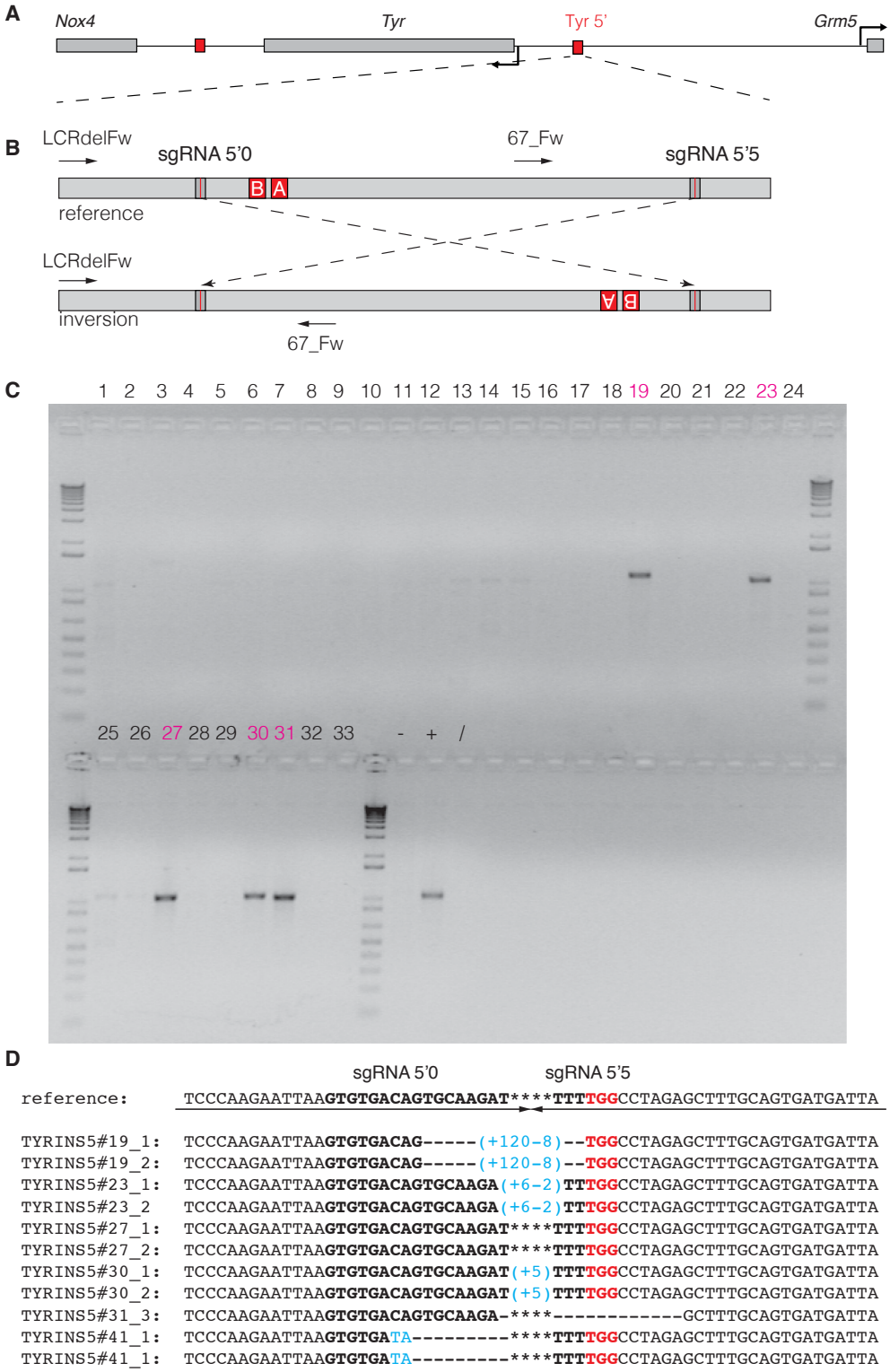


Figure r37: Chromosomal inversion of the *Tyr* 5' fragment. **A** Position of the CRISPR target sites with respect of the *Tyr* locus. **B** PCR strategy used to detect inversions: two oligonucleotides are used, LCRdelFw and 67_Fw both mapping on the same DNA strand. Only upon inversion of the DNA fragment flanked by the two CRISPR sites, the two oligonucleotides indicated yield a 1070 bp product. **C** Five individuals are positive for the inversion of the sequence flanked by the two sgRNA sites. Notably, individual #30 also carries a chromosomal deletion of the very same sequence, on the other allele. **D** Sequence alignment of two independent clone obtained from the animals in C. As observed for the deletions, also each inversion event is

unique. Inversion break-point is found in the close proximity of the sgRNA site (in bold, PAM motif in red), and insertion (in blue) and small deletions are frequently found.

4.3.9 Assessing overall mutagenesis rate

In order to evaluate the overall mutagenesis rate of each single sgRNA, we performed T7 Endonuclease I assay of all the 64 individuals obtained in the microinjection session performed using the two highest RNA concentrations, indicated in **Table r9**. These data are useful to estimate the overall mutagenesis efficiency of the CRISPR/Cas9 system and also to evaluate the frequency of double-positives in relation with the number of large deletion produced. **Figure r38 and r39** show the result of the T7 Endonuclease assay directed of the sequences flanking the sgRNA 5'0 and 5'5.

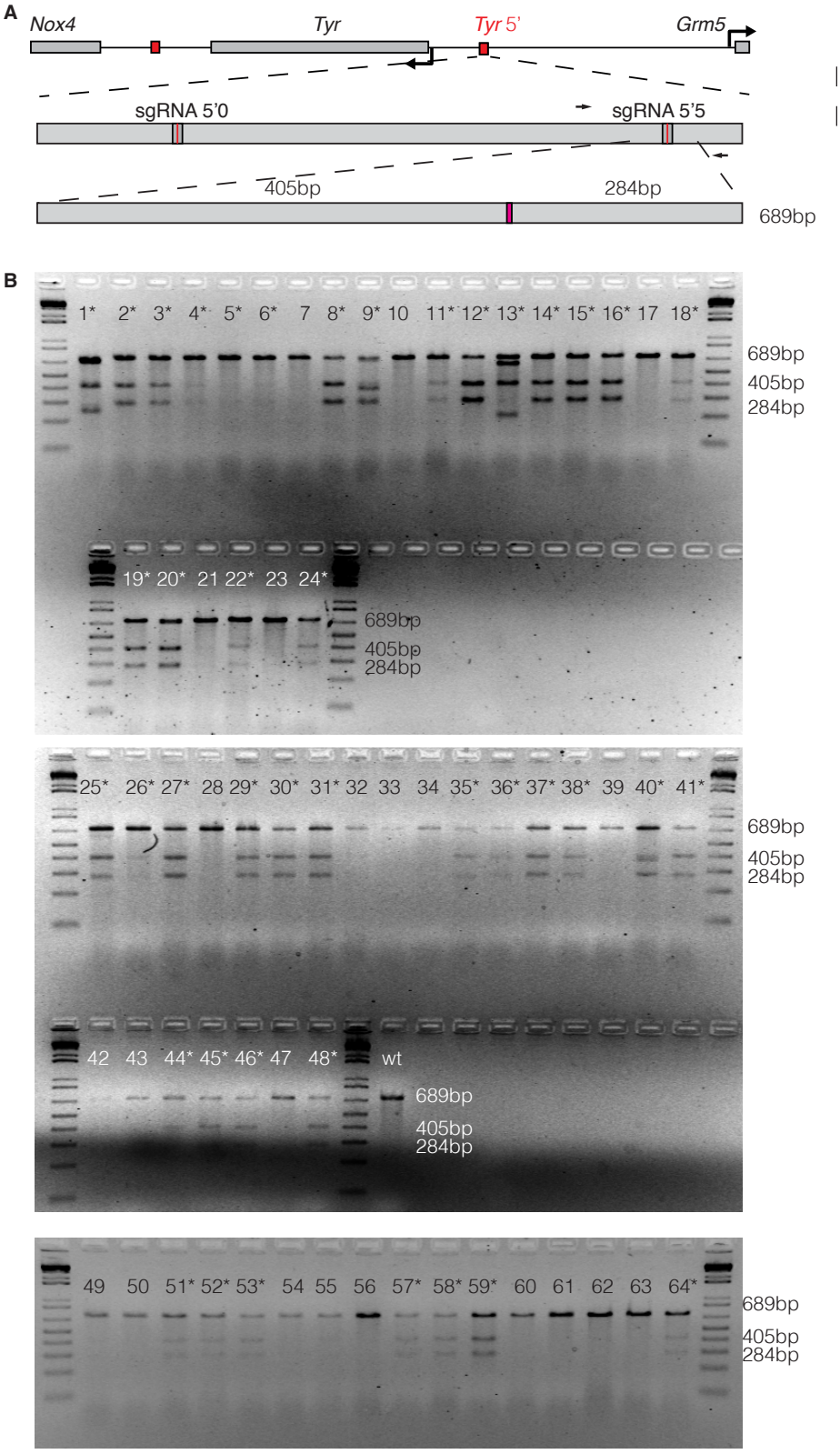


Figure r38 T7 endonuclease I assay of CRISPR 5'5. **A** Position of the CRISPR target sites with respect of the *Tyr* locus. A 689 bp sequence centred on the sgRNA 5'5 site was PCR amplified using the oligonucleotides 67Fw and LCRdelRv and subjected to T7 Endonuclease I digestion. **B** 66% of the animals analysed showed *indels* at the CRISPR 5'5 target site. The intensity of the digested bands might indicate the degree of moisaicism of the animal. Samples are scored as positive are indicated with a star.

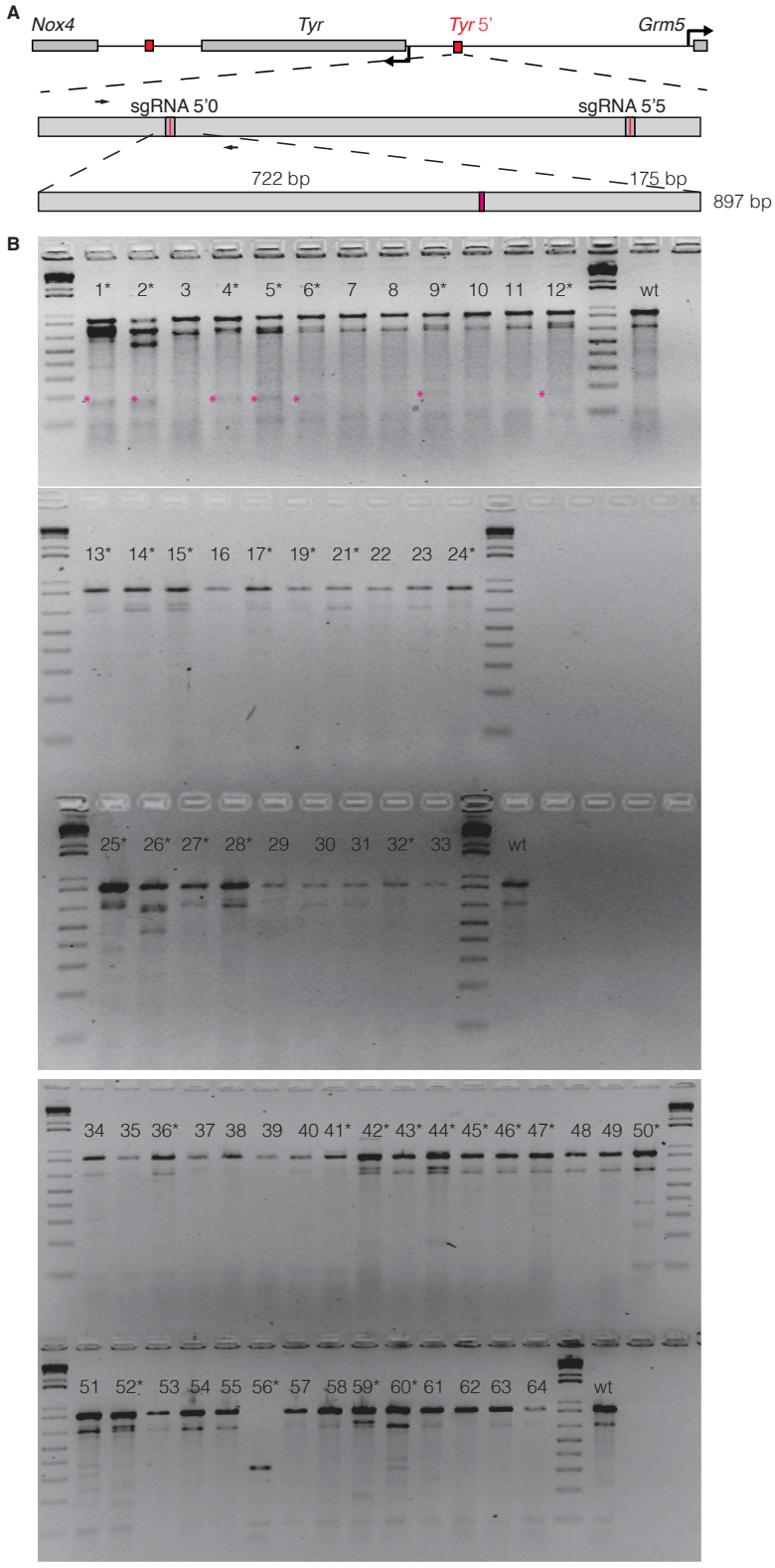


Figure r39 T7 endonuclease I assay of CRISPR 5'0. **A** Position of the CRISPR target sites with respect of the *Tyr* locus. A 897 bp sequence centred on the sgRNA 5'0 site was PCR amplified and subjected to T7 Endonuclease I digestion. **B** 50% of the animals analysed showed *indels* at the CRISPR 5'0 target site. The intensity of the digested bands might indicate the degree of moisaicism of the animal. Samples are scored as positive are indicated with a star.

The table below summarises the results obtained *in vivo*, in transgenic mice, with all the tested RNA concentrations.

Table r10: Overall mutagenesis rate

RNA concentration	positive for T7EI 5'0	positive for T7EI 5'5	positive for deletion	positive for inversion	any mutation
20 ng/µl Cas9 10 ng/µl sgRNA	1 (2%)	1 (2%)	1 (2%)	na	1
25 ng/µl Cas9 45 ng/µl sgRNA	0	0	0	na	0
35 ng/µl Cas9 60 ng/µl sgRNA	13 (41%)	17 (60%)	7 (20%)	2 (6.5%)	24 (77%)
50 ng/µl Cas9 90 ng/µl sgRNA	19 (57%)	25 (75%)	11 (33%)	5 (15%)	31 (93%)
TOTAL*	32 (50%)	42 (66%)	18 (28%)	7 (11%)	55 (86%)

* taking in account only the last two microinjection sessions

Considering the last two microinjection sessions we obtained an overall deletion frequency of 18%. If we refer only to the most successful condition, using 50 ng/ul of Cas9 mRNA, this frequency reaches the 33%. Inversions are obtained with a frequency of 11%. Frequencies of deletion, inversion and individual targeting to each of the two sgRNA sites are dose dependent. Notably, 86% of the animals carry at least a mutation. Of a total of 64 animals, we could only find nine mice that do not carry any mutation.

4.3.10 Deleting *Tyr* 3' boundary *in vivo* with CRISPRs

As done for the *Tyr* 5' boundary, *Tyr* 3' CRISPR sgRNAs were prepared for microinjection in mouse fertilized eggs (B6D2F1 background). A first microinjection session was performed with 20 ng/µl Cas9 mRNA along with 10 ng/µl of the sgRNAs 3'3 and 3'5.

Table r11 Results of CRISPR 3'3 + 3'5 microinjection

RNA concentration	Injected	Transferred	Fosters	Pups (Total - stillbirths)
20 ng/µl + 2 x 10 ng/µl	130	72 (55%)	4	14
20 ng/µl + 2 x 10 ng/µl	140	72 (51%)	4	18
TOTAL	**270**	**144 (53%)**	**8**	**32**

We genotyped the 32 animals generated from this microinjection series and we could not detect any animal carrying the desired deletion (**Figure r40**).

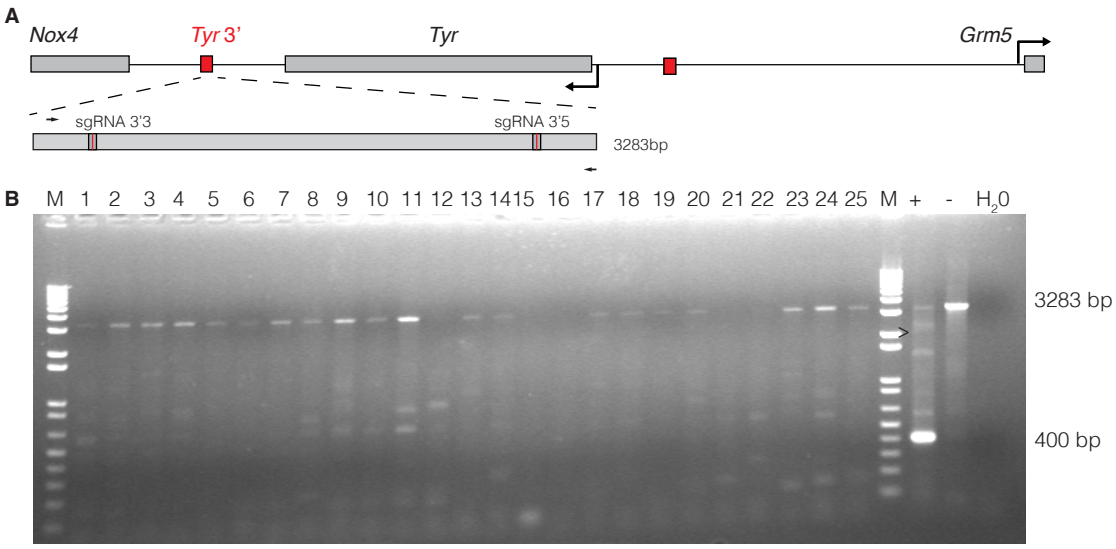


Figure r40 Results of the Tyr 3' CRISPR/Cas9 micro-injection. **A** Position of the *Tyr* 3' sgRNA sites with respect of the *Tyr* locus. A 3283 bp PCR is used for genotyping. A 420 bp PCR product is expected in case of deletion. **B** None of the animals obtained from the CRISPR microinjection resulted in the expected deletion. Genomic DNA extracted from Neuro2A cells, transfected with sgRNA 3'3, 3'5 and a Cas9 expression vector is used as positive control (+), untransfected cells are used as negative control (-).

Next, by T7 Endonuclease assay, cloning and sequencing we studied the presence of nuclease footprints at the two targeted sites. As shown in **Figure r41**, 2 out of the 16 tested individuals carry *indels* in correspondence of the CRISPR 3'3 binding site as shown by the T7 assay. Sequencing of animal #12 confirmed these findings. A 2 bp deletion and a 2 bp insertion is observed in this individual and confirmed by the sequencing of 4 independent clones (**Figure r41 A and B**).

By T7 assay of the CRISPR 3'5 target sequence, we observed digestion products in all tested animals (**Figure r41 D**), including wild-type controls. Sequencing of the CRISPR 3'5 region highlighted the presence of a microsatellite variation (**Figure r41 C**). Being in heterozygosis, this variation is detected by the T7 endonuclease I assay although it is not related with nuclease activity. By sequencing, the 3'5 CRISPR site appeared unmodified.

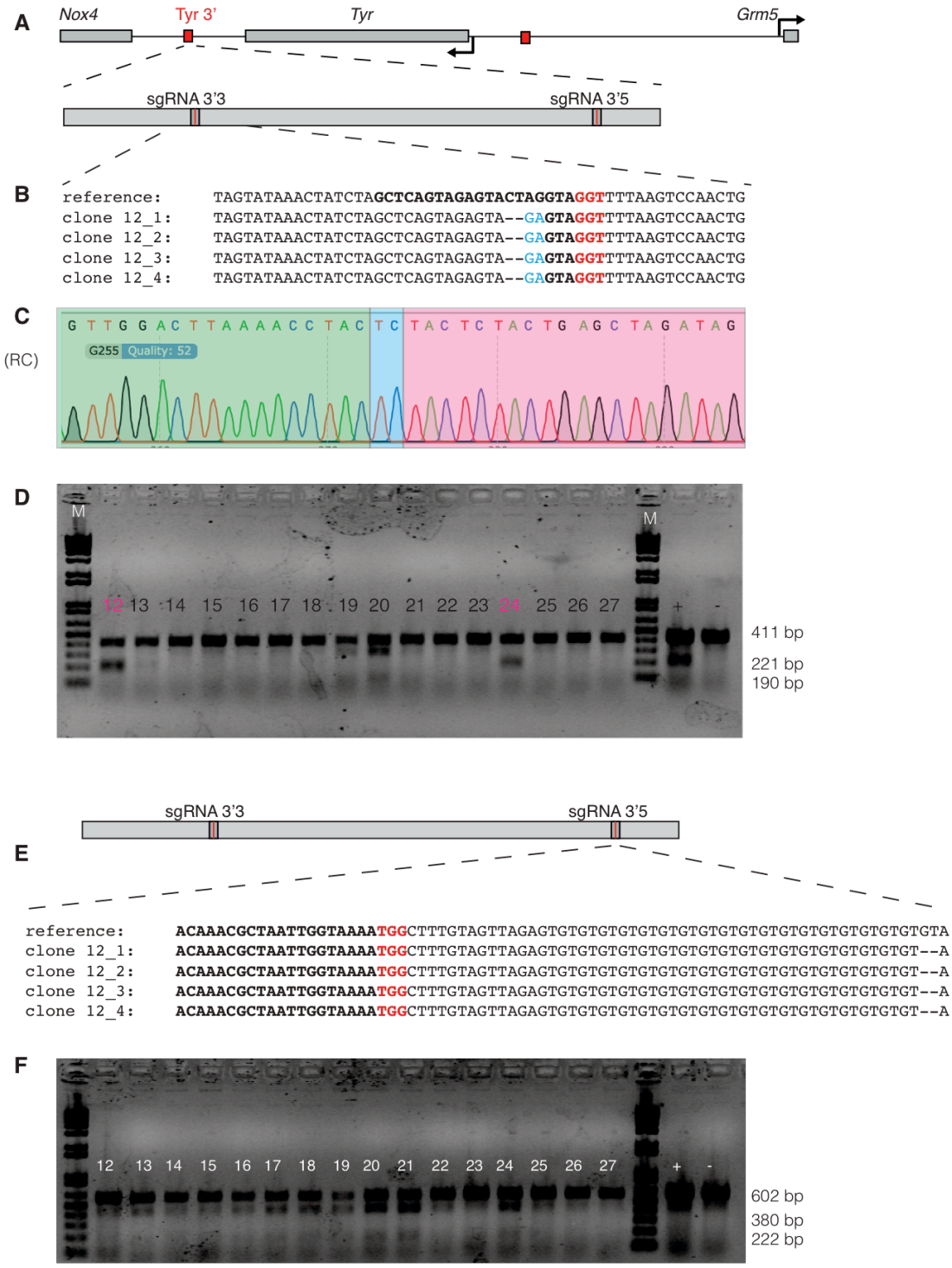


Figure r41 Sequence characterization of animal TYRINS3 #12. **A** Position of the *Tyr* 3' sgRNA sites with respect of the *Tyr* locus. **B** A PCR product centred on the sgRNA site 3'3 was cloned and sequenced. Alignments of four independent clones of animal #12 are shown. An *indel* was found 3 bp upstream the PAM motif. **C** A representative chromatogram of the sgRNA site 3'3 in animal #12 is shown. **D** Oligonucleotides 3delFw and 3T7EI123Rv were used to amplify a 411 bp sequence centred on the sgRNA 3'3 site. T7 Endonuclease assay was performed, indicating that two of the tested individuals carry *indels* at the targeted site. **E** Sequence alignment of the sgRNA 3'5 region in animal #12. The sgRNA site (in bold) is found unmodified, indicating no nuclease activity. **F** T7 Endonuclease I assay of the same region, amplified with oligonucleotides 3T7EI456Fw and Rv, did not show the digestion products corresponding to *indels* at the targeted site. Additional bands are found in all samples, including controls, that are explained by presence of

a microsatellite variation in the mouse strain used for microinjection, generating a mismatch. The microsatellite variation is also shown in all clones sequenced in **E**.

These findings suggested us that the reagents we were injecting were functioning although not at the desired efficiency. We decided to perform a second series of microinjections, as done for the *Tyr* 5' deletion, using a higher RNA dose. Results are summarized below:

RNA concentration	Injected	Transferred	Fosters	Pups (Total - stillbirths)
25 ng/µl + 2 x 45 ng/µl	300	142 (47%)	6	10
25 ng/µl + 2 x 45 ng/µl	150	78 (52%)	4	33
TOTAL	450	220 (48%)	10	43

Table r12 Results of CRISPR 3'3 + 3'5 microinjection/2

43 pups were obtained from these microinjections. Genotyping by PCR showed that 4 animals carrying a deleted allele including three different kinds of deletions (**Figure r42**)

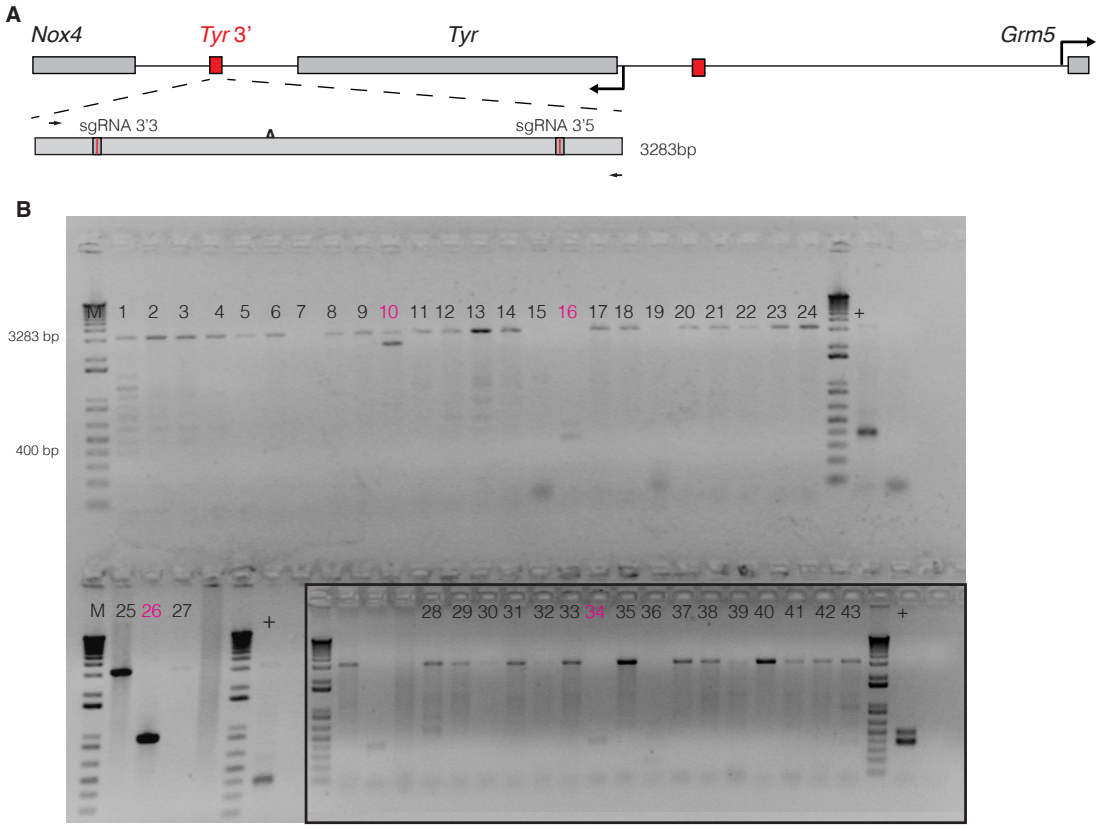


Figure r42 Results of the Tyr 3' CRISPR/Cas9 micro-injection/2. A Position of the *Tyr* 3' sgRNA sites with respect of the *Tyr* locus. A 3283 bp PCR is used for genotyping. A 420 bp PCR product is expected in case of deletion. **B** Animals #16 and 34 resulted in the expected deletion. Partial deletions are observed in animals #10 and #26. Genomic DNA extracted from Neuro2A cells, transfected with sgRNA 3'3, 3'5 and a Cas9 expression vector is used as positive control (+)

From the genotyping PCR we concluded that mouse #10 carries a partial deletion, in heterozygosis (the full-length PCR product corresponding to the wild-type sequence can be observed). Mouse #16 and #34 show the desired, full *Tyr* 3' deletion. From this PCR, it is impossible to discriminate if these mice carry one or two deleted alleles, since the small PCR product competes with the larger 3.2 kb PCR product. Mouse #26 carries an intermediate deletion. The PCR fragments of these mice were cloned and sequenced to confirm the genotype and to finely map the breakpoint of the deletions (**Figure r43 and r44**).

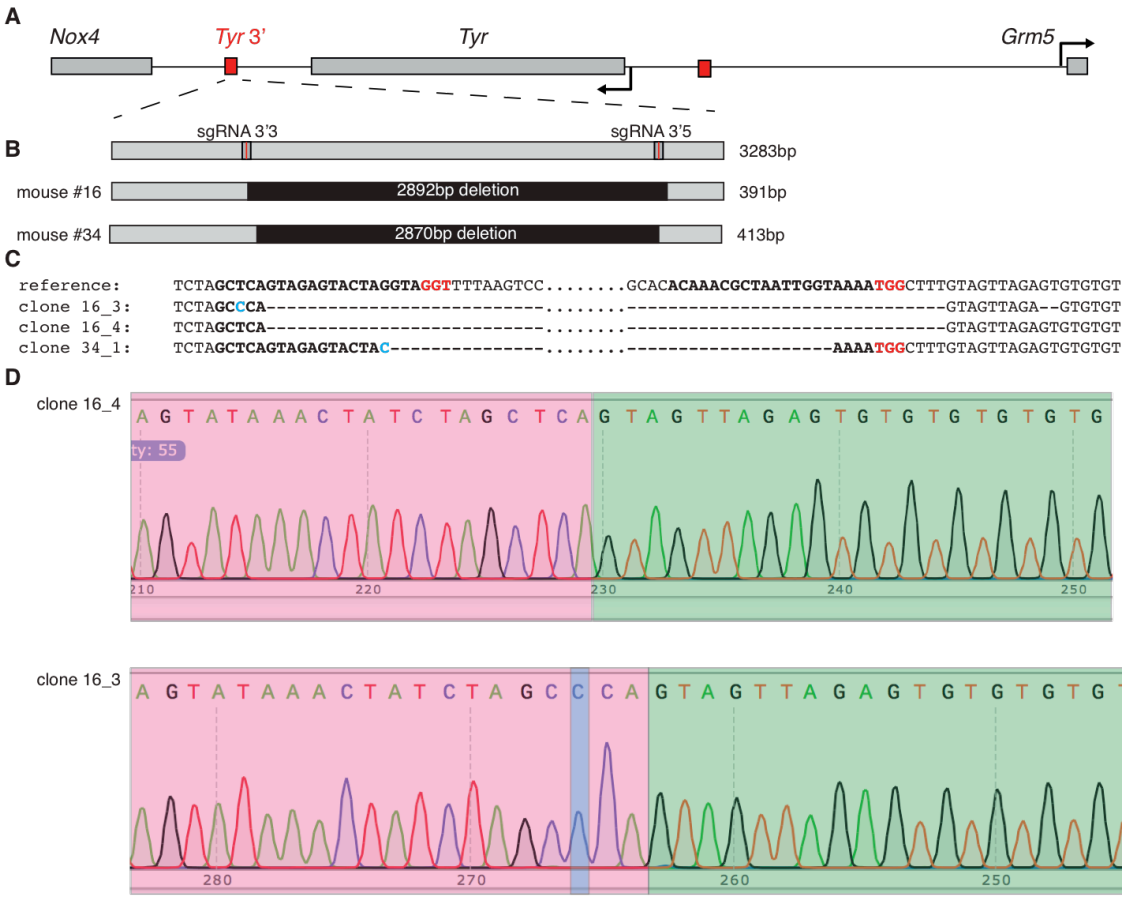


Figure r43 Sequence characterization of animals TYRINS3 #16 and #34. **A** Position of the *Tyr* 3'3 sgRNA site with respect of the *Tyr* locus. **B** Scheme illustrating the deletion produced in animals #16 and #34 with respect of the sgRNA sites. **C** Alignment of the cloning/sequencing of the deletion break-point. The sgRNA sites are shown in bold, PAM motif in red, inserted bases in blue. **D** Representative chromatograms of the break-point in animal 16.

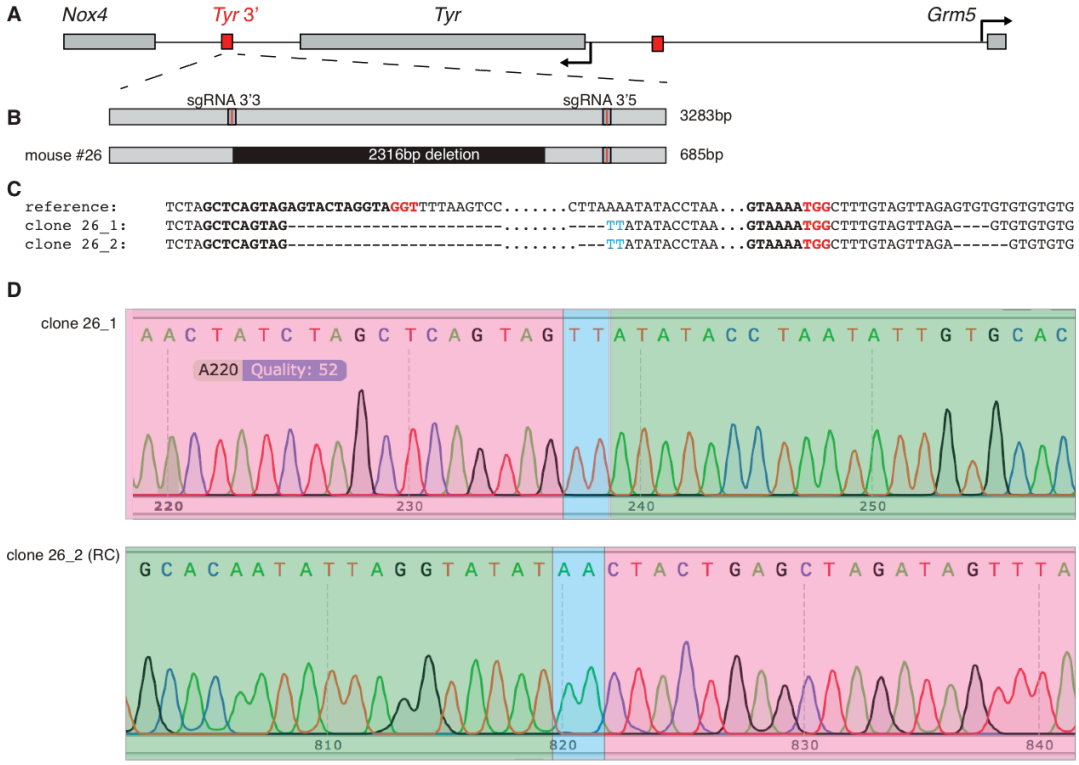


Figure r44 Sequence characterization of animals TYRINS3 #26. **A** Position of the *Tyr* 3'3 sgRNA site with respect of the *Tyr* locus. **B** Scheme illustrating the deletion produced in animals #26 with respect of the sgRNA sites. **C** Alignment of the cloning/sequencing: the 3' break point is found 565 bp upstream the expected CRISPR 3'5 cleavage site. The sgRNA sites are shown in bold, PAM motif in red, inserted bases in blue. **D** Representative chromatograms of the break-point in animal 26. The two distal sequences are highlighted in green and red. *Indels* are highlighted in blue.

The core of the *Tyr* 3' boundary, composed by a SINEb2 element and a CTCF-binding site, was successfully deleted in both TYRINS3#16 and TYRINS3#26.

In conclusion, we obtained different mutated alleles of the *Tyr* 3' boundary region, that are summarized in **Figure r45**.

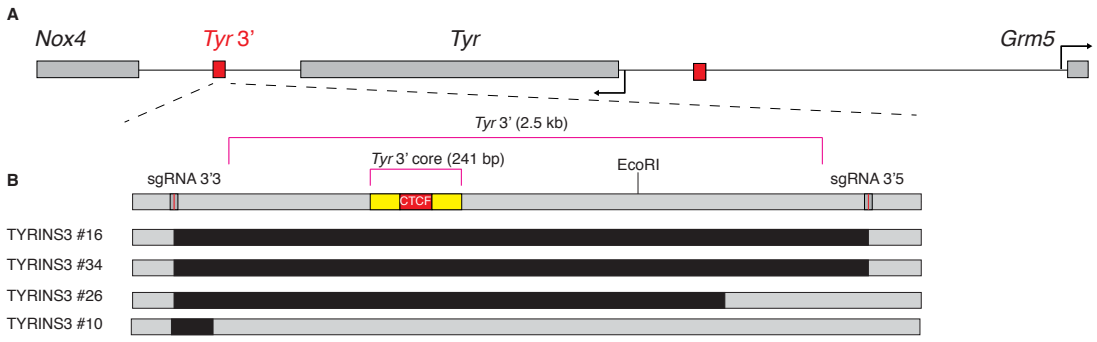


Figure r45 Founder individuals of the TYRINS3 series lines **A** Position of the sgRNA sites with respect of the *Tyr* locus. **B** Scheme of the deletions produced *in vivo* with CRISPR. All the sequences associated with regulatory features analysed in this work, including a *Dpn*II fragment interacting with the *Tyr* promoter, a 2.5 kb insulator sequence and its core element, 271 bp large, including a SINEB2 element and a CTCF-binding site, are deleted in animals #16, 26 and 34, but not in #10.

4.3.11 Germline transmission

Selected individuals from the cohort of TYRINS5 founders were bred with albino outbred HsdWin:NMRI wild-type mice. All individuals carrying a deletion and presenting alterations in coat colour were considered for breeding. Additionally, animals carrying a partial deletion (individual #8) and larger deletion (#13) were bred. Animals carrying inversions and a combination of deletion and inversion were also selected for breeding. **Table r13** illustrates the germ-lime transmission of the tested founders.

Founder	feature	GLT	Transgenic/total
TYRINS5#8	Partial deletion	YES	10/24
TYRINS5#11	Full deletion	YES	1/19
TYRINS5#13	Larger deletion	YES	5/15
TYRINS5#18	Full deletion Partial deletion	YES YES	6/9 3/9
TYRINS5#19	Inversion	YES	1/7
TYRINS5#23	Inversion	NO	0/2
TYRINS5#30	Deletion Inversion	YES NO	2/12 0/12
TYRINS5#31	Inversion	NO	0/27
TYRINS5#35	Full deletion	YES	3/21
TYRINS5#41	Inversion	YES	7/17
TYRINS5#50	Full deletion	YES	7/13
TYRINS5#60	Deletion Inversion	YES NO	8/21 0/21
TYRINS5#40	Partial deletion	NO	0/14

Tabel r13 Germline transmission of selected *Tyr* 5' series founder mice.

Founder TYRINS5#18 carries a 1.2 kb deletion, plus a second smaller deletion that we could not verify by cloning and sequencing in the founder animal (**Figure 35**), probably because the smaller band competes with the larger one in the PCR reaction as well as in the cloning reaction. Upon segregation of the two different alleles in the TYRINS5#18 F1 generation, we could indeed detect this additional band, that is now confirmed as a second, smaller deletion present in the founder and transmitted to three of the nine F1 individuals (**Figure 46**). All the lines carrying a deletion transmitted to the germline. The ratio of GLT for each founder may reflect different degrees of mosaicism. Transmission of the inverted TYRINS5 allele is problematic: out of six founders (including male and female founders) tested for GLT, only two (TYRINS5#19 and #41) transmitted the genetic modification to the F1 generation.

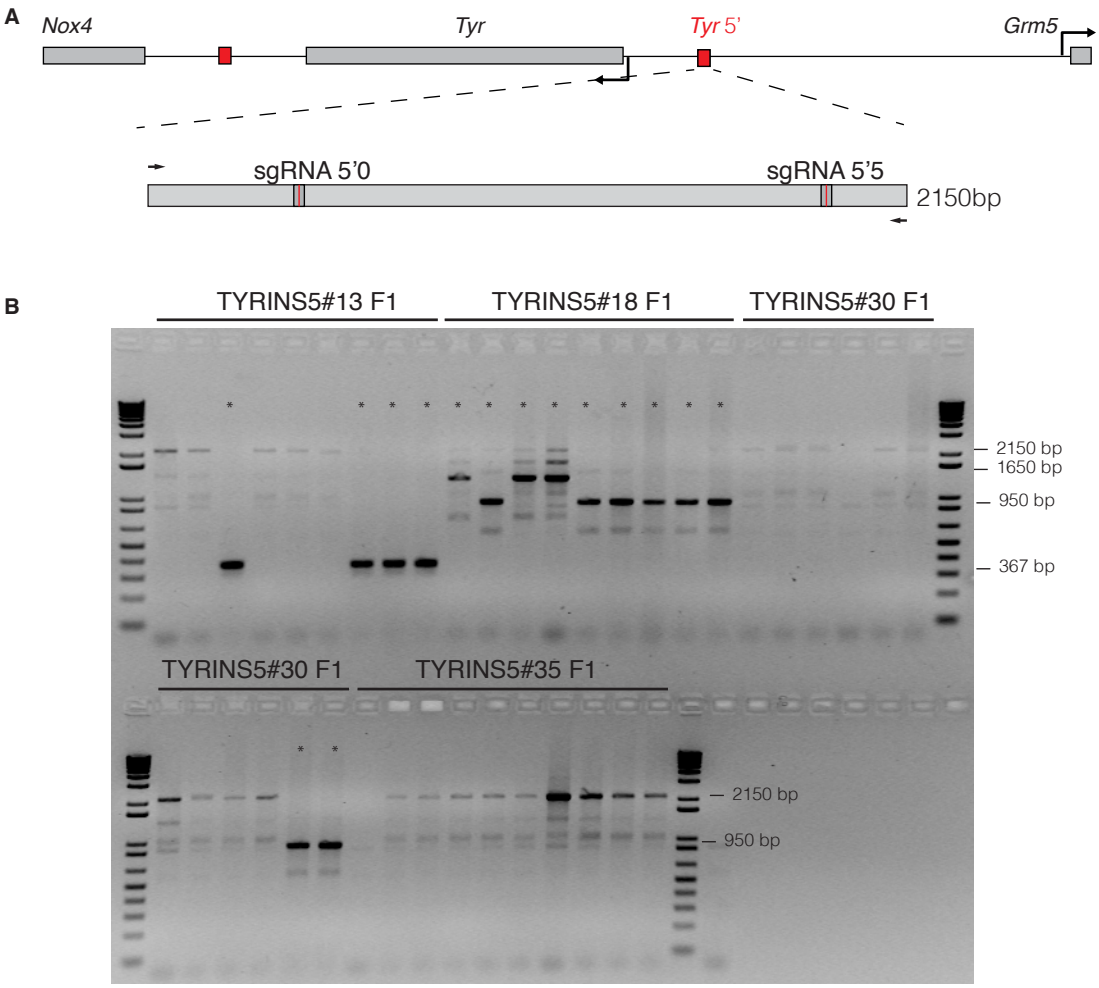


Figure 46 Genotyping of TYRINS5-series F1 individuals. **A** Position of the *Tyr* 5' element with respect of the *Tyr* locus. A 2150 bp band corresponds to the wild-type allele. Smaller PCR product are obtained according to the size of the deletion. **B** Genotyping of the F1 individuals from lines TYRINS5#13, #18, #30 and #35. Note that two different deletions segregated in the TYRINS5#18 line. Positive samples are indicated with a star.

Two individuals carrying full and partial deletion of the targeted *Tyr* 3' (TYRINS3 #16 and #26, both female) were crossed with C57B6/J males. Analysis of the resulting F1 animals confirmed the germ-line transmission of the genetic modifications (**Table r14**).

Table r14 Germline transmission of selected *Tyr* 3' series founder mice.

Founder	feature	GLT	Transgenic/total
TYRINS3#16	Full deletion	YES	4/6
TYRINS3#26	Partial deletion	YES	1/7

4.3.12 Analysis of predicted CRISPR off-targets

Several independent laboratories reported that the CRISPR/Cas9 system might be able to induce mutation at chromosomal loci with high sequence similarity with the intended target (Fu et al. 2013; Mali et al. 2013; Pattanayak et al. 2013; Hsu et al. 2013). **Table r15** illustrates the sequences of three top-scoring predicted off-targets for sgRNA 5'0 and 5'5, that were used to delete the *Tyr* 5' boundary. **Table r16** shows the top three predicted off-target sites for the sgRNA used to delete the *Tyr* 3' boundary element.

Taking this observation in account, we analysed the presence of mutations at the three most similar sequences for each of the sgRNA sequence that were used to delete the *Tyr* 5' element.

Table r15 Predicted off-targets of sgRNA 5'0 and 5'5. The top three off-target (OT) sites for each sgRNA used are shown. Mismatching bases are shown in red. All off-target sites map on intronic or intergenic, non-coding region. None of the predicted off-targets maps on chromosome 7.

ID	sequence	chromosomal position	feature
sgRNA 5'0	GTGTGACAGTGCAAGATAAC-AGG	chr7: 94656546	intergenic
5'0-OT1	CTGTCACAGAACAAGATAAC-CGG	chr6: 99084014	intronic (*Foxp1*)
5'0-OT2	GGGAGACTGTGCAAGATAAG-GGG	chrX: 69682563	intronic (*Gabra3*)
5'0-OT3	TTTTGACAGAGCAAGATAAA-TGG	chr3: 60502350	intergenic
sgRNA 5'5	GTGATAAAACTAGGCAATTT-TGG	chr7: 94657775	intergenic
5'5-OT1	CAGATAAAAGCAGGCAATTT-TGG	chr1: 9300003	intergenic
5'5-OT2	GAGCAAAAACTTGGCAATTT-GGG	chr19: 31071409	intronic (*Prkg1*)
5'5-OT3	GTATTAAAATGAGGCAATTT-GGG	chr14: 112763119	intergenic

Table r16 Predicted off-targets of sgRNA 3'3 and 3'5. The top three off-target (OT) sites for each sgRNA used are shown. Mismatching bases are shown in red. All off-target sites map on intronic or intergenic, non-coding region. None of the predicted off-targets maps on chromosome 7.

ID	sequence	chromosomal position	feature
sgRNA 3'3	AGCTCAGTAGAGTACTAGGT-AGG	chr7: 94656546	intergenic
3'3-OT1	AGCTGGGCAGAGTACTAGGT-TGG	chr11: 74783860	intronic (*Smg6*)
3'3-OT2	CTCTGAGTAAAGTACTAGGT-TGG	chr11: 17927436	intergenic
3'3-OT3	AGCTCAGGAGAGTCCTAGGT-AGG	chr5: 135521999	intronic (*Stx1a*)
sgRNA 3'5	ACAAACGCTAATTGGTAAAA-TGG	chr7: 94657775	intergenic
3'5-OT1	ACAATTGGTAATTGGTAAAA-TGG	chr12: 36137724	intergenic
3'5-OT2	AGATTCGCTGATTGGTAAAA-AGG	chr5: 42142945	intergenic
3'5-OT3	ACAAATAGTAATTGGTAAAA-AGG	chr15: 15417522	intergenic

By T7 Endonuclease assay, we scanned the presence of mutations in all the 64 individuals obtained in the last two series of microinjection. We analysed positive individuals as well as littermates not bearing the desired deletions. In total, we performed six series of T7 EndonucleaseI assays each one involving the 64 founders plus a number of control animals of the same genetic background. With few exceptions (illustrated below) we could not detect mutations at the six off-target sites in anaysis. Regarding the off-target site 5'0 OT1, located in the intronic region of *Foxp1*, individuals #13, #14, #29 (**Figure r46 B**) and #57 showed faint digested bands at the T7 Endonuclease I assay. However, sequencing of three independent clones did not confirmed the presence of off-target mutations (**Figure r46 C**), probably indicating that the mutations detected were present in a very small number of cells in the sampled biopsies.

Similarly, by T7 Endonuclease I assay we detected the presence of a mutation at the first predicted off-target for the 5'5 sgRNA sequence (55 OT1) in two: #1 and #24. Also in this case, by sequencing three independent clones, no mutation could be detected (**Figure r47 D**).

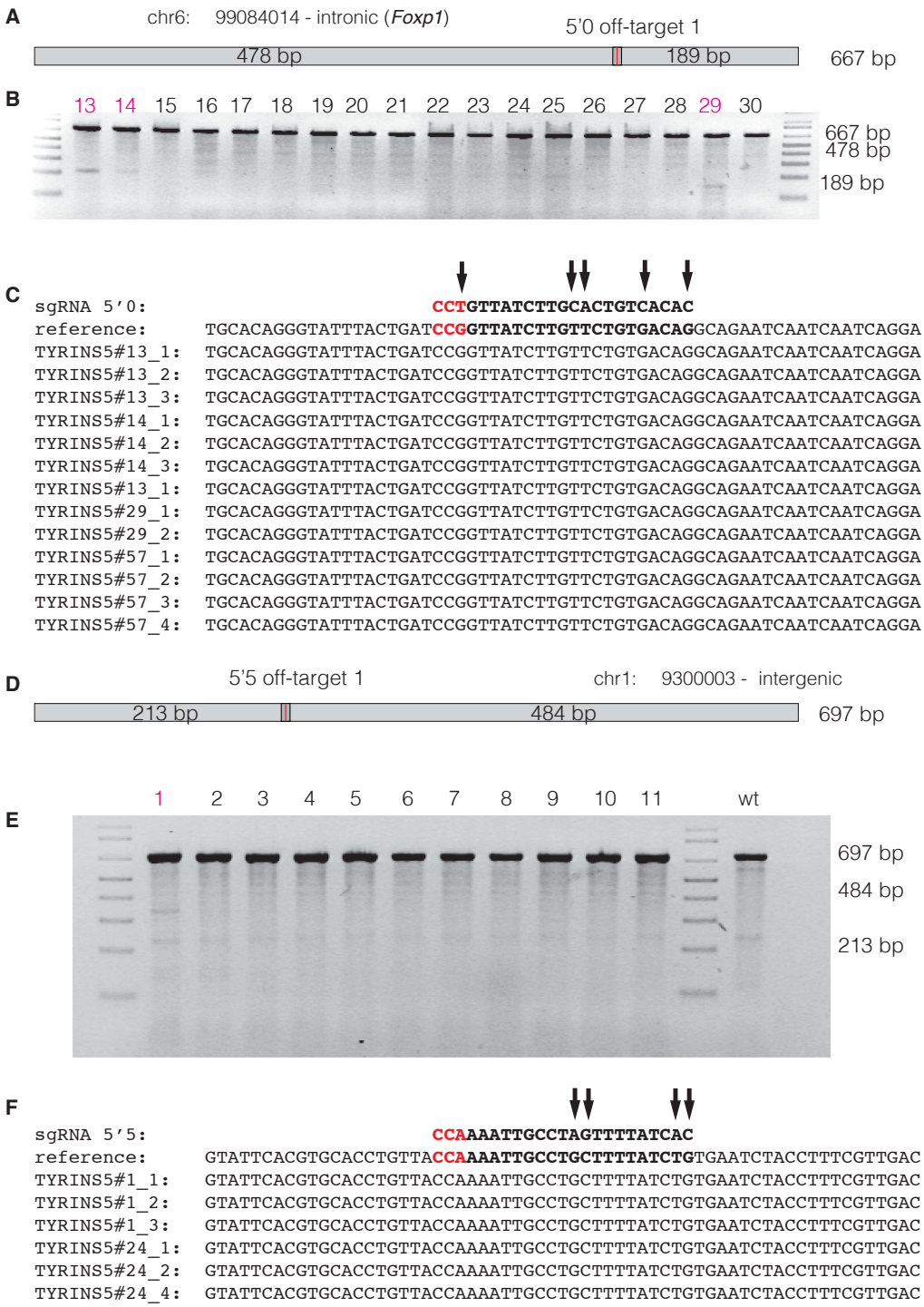


Figure r47 Analysis of off-target sites 5'0 OT1 and 5'5 OT1. **A** Position of the off-target site 5'0 OT1. **B** Analysis of the presence of *indels* by T7 Endonuclease I assay. Faint bands were obtained in samples #13, 14 and 29. The size of the observed bands does not match with the pattern expected upon off-target mutation **C** The sgRNA 5'0 site is aligned with the sequence of its predicted off-target. Black arrows indicate mismatches. By cloning and sequences, we could not identify any mutation induced at the off-target site 5'0-OT1. **D** Position of the off-target site 5'5 OT1. **E** Similarly, the 5'5 OT1 off-target site was tested by T7 Endonuclease I assay. Faint bands were observed in samples #1 and #24 (the latter not shown). **F** By cloning and Sanger sequencing, no mutation at the sgRNA 5'5 OT1 off-target site were found in individuals previously screened by T7 Endonuclease assay.

4.4 Characterization of the TYRINS5 and TYRINS3-series of transgenic mouse lines

4.4.1 Phenotype of TYRINS5-series founder animals

Previous experiments in our laboratory showed that YAC constructs of the mouse *Tyr* locus lacking the *Tyr* 5' insulator (3.7 kb, LCR) failed to fully rescue the albino phenotype when injected in albino fertilized eggs (**Figure i6**; Montoliu et al. 1996). Additionally, YACs carrying the specific deletion of the A box alone or of A plus B boxes, showed a partial rescue of pigmentation, indicating that these sequences are indeed relevant for *Tyr* expression. However, the degree of pigmentation in transgenic mouse lines generated with YRT2ΔA and YRT2ΔAB YACs varies between transgenic lines. (**Figure i7**; Giraldo and Montoliu, 2002; PhD thesis of Patricia Giraldo, 2002).

By deleting a 1.2 kb fragment of the *Tyr* LCR in the endogenous locus, we obtained – out of the eighteen positive animals generated – four individual founder mice bearing evident reduction in coat colour pigmentation. Considering that *Tyr* is a dominant allele, chromosomal deletion must have happened at both alleles. This phenotype suggests the relevance of the *Tyr* 5' element within its natural, chromosomal location and suggests its potential implication on a mutation associated to OCAI (King et al. 2003; Regales et al. 2003; Montoliu et al. 2014).

In these four founder animals (**Figure r48**) both copies of the target element could be deleted in a consistent number of cells. The degree of pigmentation is variable between the four founder animals, both in term of intensity and spatial distribution, indicating different degree of mosaicism, likely caused by the time during embryonic development in which these deletion took place. Animal #18 shows the strongest phenotype, hence the weakest pigmentation, suggesting that the *Tyr* 5' element was deleted very early in embryogenesis and, thus, this element is completely absent in most cells of the mouse body. Next, animal #11 shows a darker coat colour, with grey patches. This suggests that the deletion of the *Tyr* 5' element in both alleles could have happened later in embryogenesis. Finally, Individuals #35 and #60 show only spare patches of grey coat, suggesting a very little content in biallelic deleted-cells.

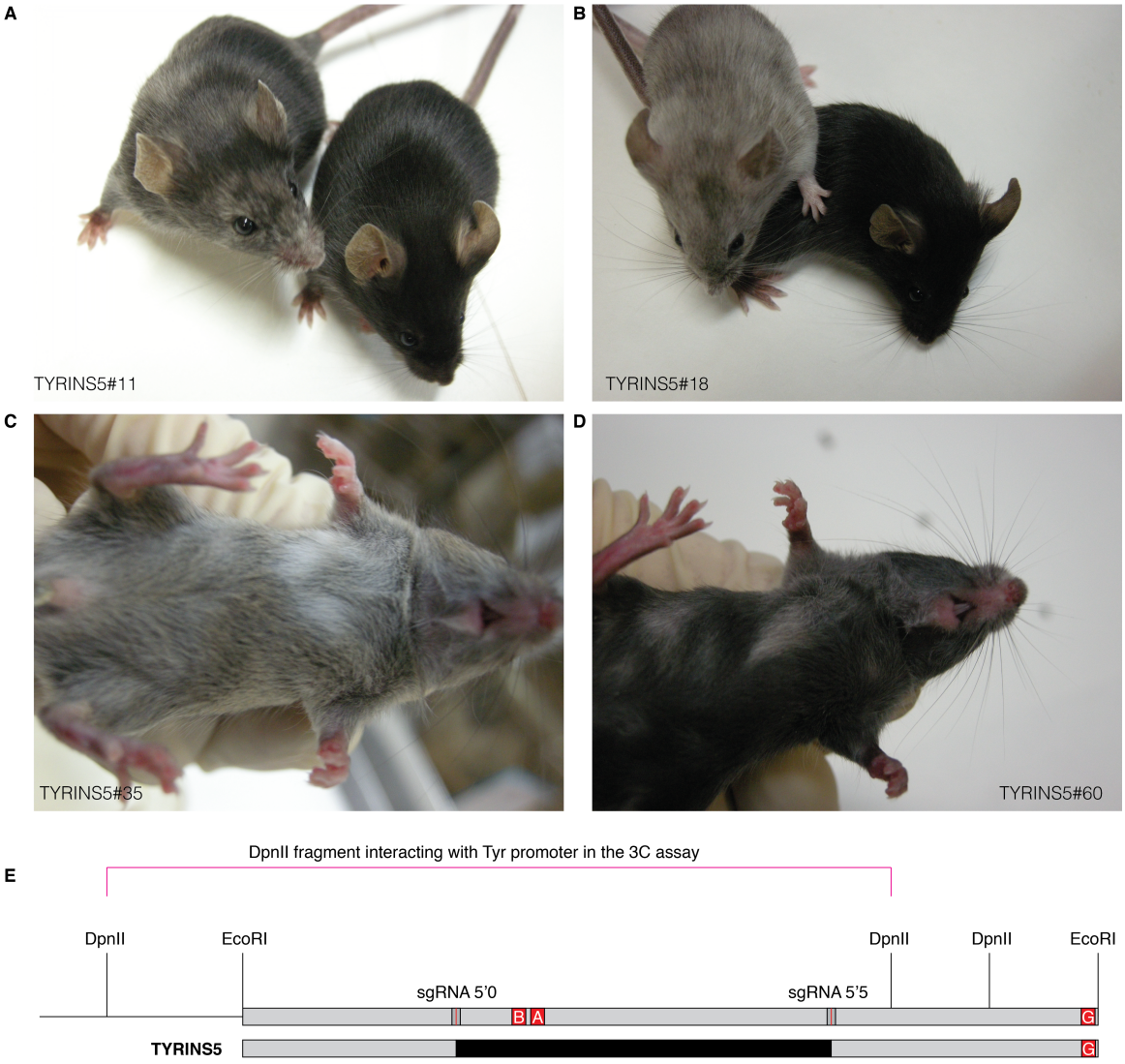


Figure r48 Alteration in coat color in F0 individuals. **A-D** Four individual obtained from CRISPR mRNA embryo microinjection show reduction in coat colour pigmentation. The reduction is more evident in animals TYRINS5#11 (in **A**, compared with a non-transgenic littermate) and #18 (**B**, compared with a non-transgenic littermate). Alterations in animals #35 (**C**) and #60 (**D**) are less evident and could only be observed in the ventral part on the animal, during the manipulation at the weaning time. **E** Schematic of the deletion present in mice shown in **A-D**.

4.4.2 Phenotype of distinct TYRINS5 transgenic lines

Selected individuals from the large cohort of founder animals (22 animals carrying either a deletion or inversion or both) were bred with albino, outbred HsdWin:NMRI wild-type mice, the genetic albino background were all the YAC-*Tyr*-based transgenic lines were generated and currently maintained. NMRI outbred mouse strain carries the classical homozygous mutation in the *Tyr* coding sequence (C103S, $Tyr^{c/c}$) resulting in albino phenotype (Jackson and Bennet 1990). In an albino background, alterations in *Tyr* expression will be evident even if only one copy of the *Tyr* 5' element is inactivated.

F1 pups that inherited the deletion, show a grey coat and pigmented eyes, but still darker than the albino coat colour (**Figure r49 A**). As indicated before, the TYRINS5#18 founder bears two distinct deletions that segregated independently in the F1 generation, named TYRINS5#18a (full deletion) and TYRINS5#18b (partial deletion). Interestingly, all F1 individuals from the TYRINS5#18 founder have the same grey coat colour, indicating that both type of deletions, are sufficient to inactivate key sequences necessary for proper *Tyr* expression (**Figure r49 B**). Of the F1 generation of the line TYRINS5#30, obtained from a founder carrying a combination of deletion and inversion, only two individuals – that inherited the deleted allele – have grey coat (**Figure r49 C**). Finally, the TYRINS5#50 F1 that inherited a deleted copy of the *Tyr* 5' sequence show the same phenotype observed in lines #8, #13, #18 and #30. In a second line that carries both inversion and deletion, line TYRINS5#60, a grey coat is observed in animals carrying a deletion.

By comparing the coat colour phenotype of lines #18a, #30, #50 and #60 with that of line #13, we can conclude that the DNA region 5' upstream to the sgRNA 5'0, deleted only in line #13, are not relevant to the regulation of *Tyr* expression. By comparing again lines #18a, #30, #50 and #60 with line #8, carrying a 0.7 kb deletion centred on the AB box, we can conclude that all additional sequences found upstream to the 3' end on the deletion in line #8 are not relevant to the proper *Tyr* expression. In fact, different lines show the same coat colour in the presence and in the absence of these sequences. Finally, by comparing lines #8 and #18b, we could trim the size of the *Tyr* 5' fragment that, when deleted, is able to alter *Tyr* expression. The lack of a 410 bp sequence, included between the sgRNA 5'0 site and the 3' break-point of the deletion in line #18b, is sufficient to alterate *Tyr* expression *in vivo*.

Of the two lines carrying the inversion and successfully transmitting to the F1 generation, both display a grey coat colour (**Figure r49 D**). This indicates that inverting a regulatory element (and particularly an insulator) reduces the functioning of the regulatory circuit.

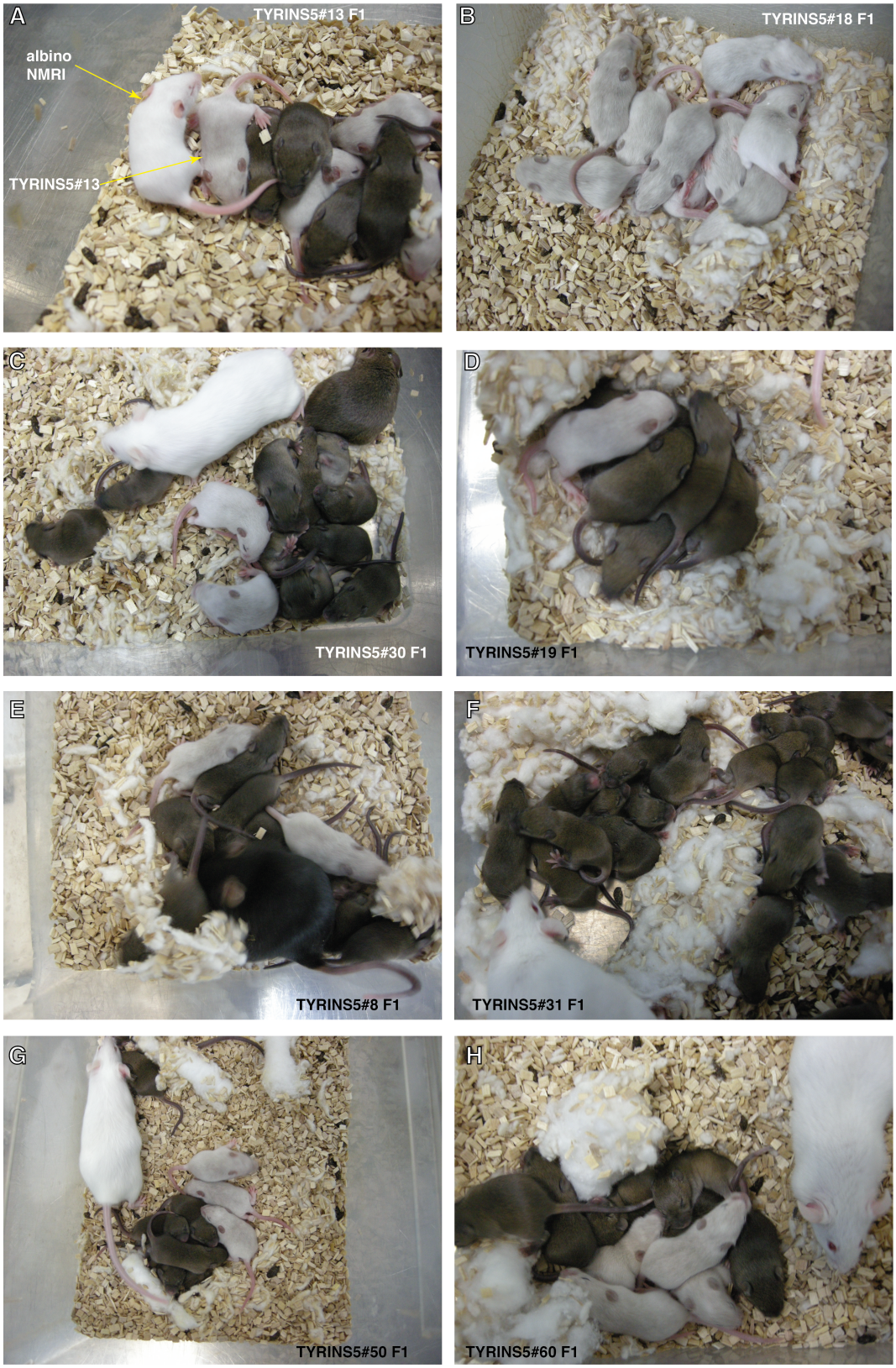


Figure r49 Phenotype of TYRINS5 lines F1 offspring. **A** Four out of nine pups of the TYRINS5#13 F1 offspring carry a 1.7 kb deletion of the *Tyr* 5' element and show a grey coat colour and pigmented eyes. As comparison, an age-matched, albino Hsd:Win wild-type animal is added aside. **B** All nine pups from the TYRINS5#18 F1 offspring show a grey coat colour and pigmented eyes, indistinguishable from the phenotype of the TYRINS5#13 line. Six of the nine TYRINS5#18 F1 individuals carry a 1.2 kb deletion, while three inherited a smaller deletion that was also produced in the founder. **C** The TYRINS5#30 founder carries a deleted and an inverted allele. Only those the two F1 individuals that inherited the deleted allele show a grey coat colour and pigmented eyes as seen in TYRINS5#13 and #18 F1s. The F1s that inherited the

inverted *Tyr* 5' sequence do not show evident alterations in coat colour pigmentation. **D** F1 individuals from the TYRINS#19 offspring. Only one mouse inherited the inverted allele and shows a grey coat similar to the TYRINS5 lines carrying a deletion of the same sequence. **E** Two out of the nine TYRINS5#8 F1 offspring show the same grey coat colour as seen for the other lines carrying the inactivation of the *Tyr* 5' element. **F** None of theTYINS5#31 F1 offspring inherited the inversion. The genotype matches with the lack of germline transmission of the inversion. **G** F1 offspring of the TYRINS5#50 lines. Three individuals show the same grey coat phenotype as observed for the previous lines. **H** The same phenotype can be observed in four individuals of the TYRINS5#60 F1 offspring, the same individuals that inherited the deleted allele from the founder.

Line	feature	With alterations in coat colour
TYRINS5#8	Partial deletion	2/2
TYRINS5#11	Deletion	1/1
TYRINS5#13	Larger deletion	4/4
TYRINS5#18/a	Deletion	6/6
TYRINS5#18/b	Partial deletion	3/3
TYRINS5#19	Inversion	1/1
TYRINS5#30	Deletion	2/2
TYRINS5#35	Deletion	3/3
TYRINS5#41	Inversion	7/7
TYRINS5#50	Full deletion	3/3
TYRINS5#60	Deletion	4/4

Table r17: Coat color phenotype of *Tyr* 5' F1 offspring.

At the histological level, the grey coat colour observed in the TYRINS5 lines corresponds to a reduced pigment accumulation in the hair follicle (**Figure r50**).

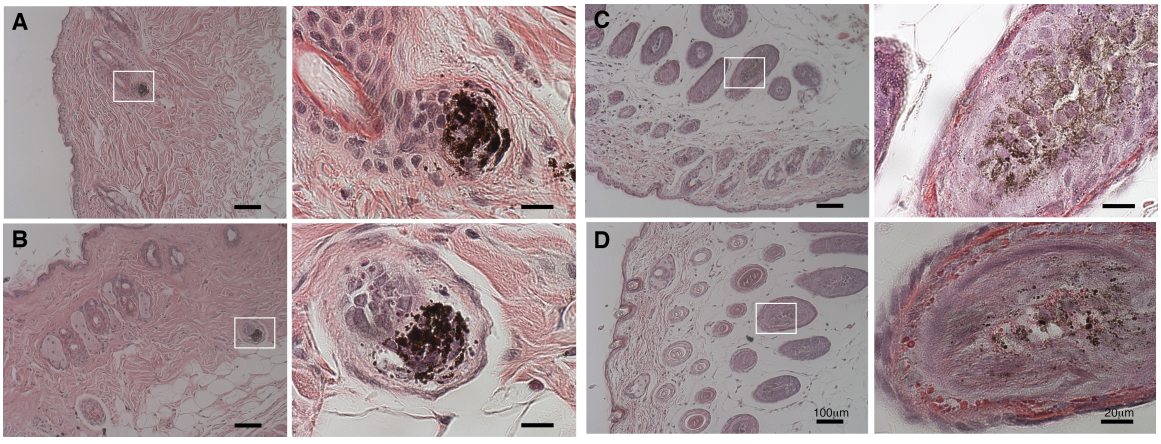


Figure r50 Skin histological sections. **A** Back skin section from a YRT2 transgenic mouse. Melanin is accumulated in the base of the hair follicle. **B** Similarly, pigmented hair follicles can be observed in the skin from a YRT3 mouse. **C** and **D** Skin sections from two individuals of the TYRINS5#18a line. Less dense melanin accumulation is observed, causing the coat colour to be grey. Scale bars 100 µm and 20 µm in insets. Hematoxylin-eosin staining.

At macroscopical observation, the eyes of all TYRINS5 lines appears pigmented and indistinguishable from wild-type mice. However, by histological analysis, a difference is observed. In the absence of the *Tyr* 5' element, *Tyr* expression is

dramatically reduced in the choroid melanocytes and but not in the RPE cells (**Figure r51**).

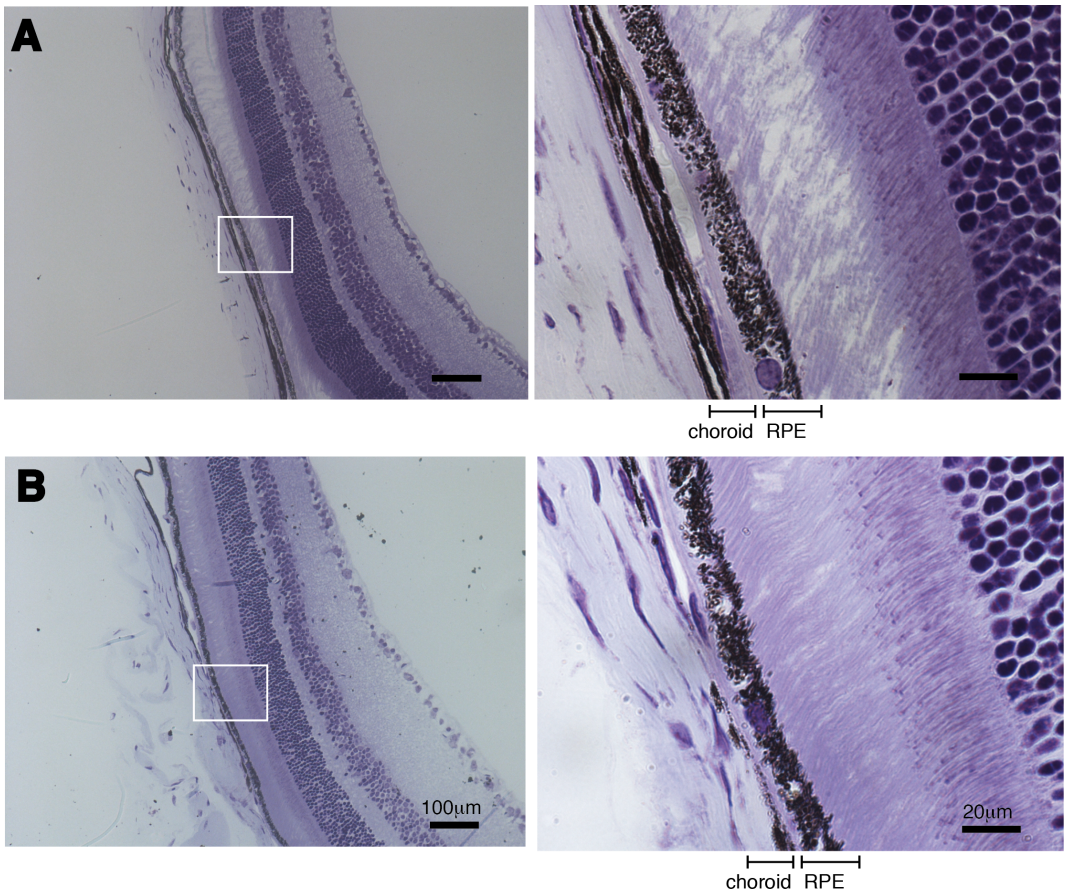


Figure r51 Eye histological sections. **A** Section from a pigmented, agouti wild-type mouse. At higher magnification, two distinct pigmented layers can be observed: the choroid and the RPE cell layer. **B** Eye section of a TYRINS5#18 mouse. Pigmentation of the choroid is almost completely lost but not in the RPE layer.

4.4.3 Phenotype of TYRINS3 series founder animals

In contrast of what had been observed with the TYRINS5 series of microinjections, no biallelic deletions were obtained in TYRINS3 F0 individuals. None of the founder individuals carrying deletion of the *Tyr* 3' element showed evident alternations in coat colour or eye pigmentation.

4.4.4 Phenotype of distinct TYRINS3 transgenic lines

TYRINS3 F0 animals (obtained in the pigmented genetic background B6D2F1) were crossed with C57B6/J individuals in the laboratory of Dr. Pawel Pelczar, University of Zurich, Switzerland. In such pigmented genetic background, coat colour alteration could not be observed. After being imported at the CNB animal

house, F1 individuals were crossed with albino outbred HsdWin:NMRI wild-type animals, as it was done with the TYRINS5 project founders. TYRINS3 F2 individuals, now in albino background, do not show coat colour alterations, suggesting that the lack of the *Tyr* 3' element does not affect *Tyr* expression (**Figure r52**).

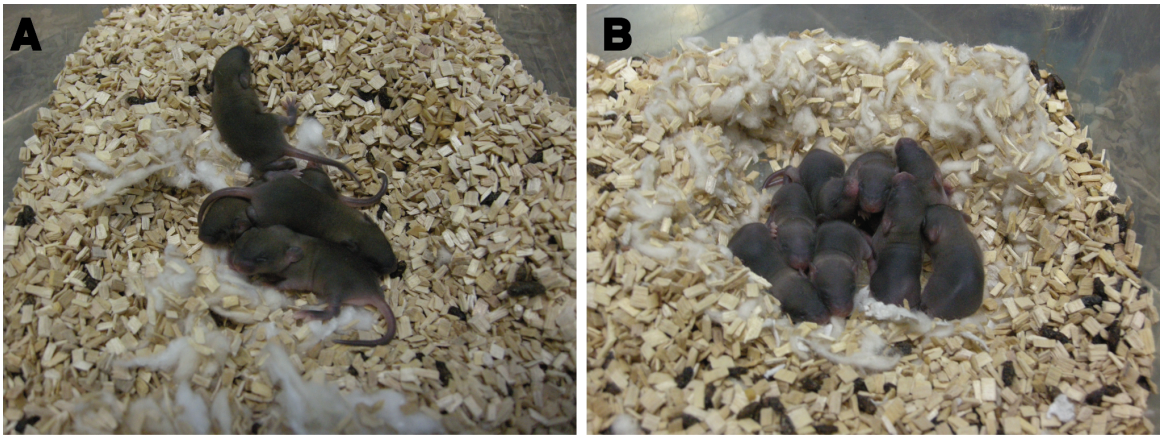


Figure r52 Phenotype of TYRINS3 lines F2 offspring. **A** TYRINS3#16 F2 offspring in albino background: normal coat colour is observed. **B** TYRINS3#26 F2 offspring in albino background show normal pigmented phenotype.

5. Discussion

5.1 The *Tyr* locus is flanked by chromatin insulators

The expression pattern of the flanking genes *Nox4* and *Gmr5* (**Figure r3**), and the very specific expression of *Tyr*, suggests that the latter needs to be ''protected'' from the nearby regulatory signals belonging to the *Nox4* and *Grm5* expression units. Indeed, both 5' and 3' intergenic regions contain sequences able to block enhancer activity when tested *in vitro* (**Figure r6**) *in vivo* (**Figure r7** and Bessa et al. 2009). The *Tyr* locus thus contains a complete expression unit flanked at either side by chromatin boundaries and containing all regulatory elements necessary to coordinate its expression in the relevant tissues and developmental stages.

The activity of the *Tyr* 3' boundary is largely depending on CTCF occupancy. After mutagenesis of the CTCF binding site, we observed a significant drop in enhancer-binding activity, but notpub a full loss of insulator capability. The residual insulator activity detected after mutagenesis of the CTCF binding site (**Figure r7 and r8**) could be due to the transcriptional activity of polymerase II and III promoters contained in the SINEB2 repeat, which is, *per se*, an element capable of transducing insulator properties (Lunyak et al. 2007; Lunyak, 2008). Another hypothesis is that with the mutagenesis strategy we used didn't abolish but only reduced the binding of CTCF. We designed the mutation following a strategy that supposedly allowed for complete loss of protein binding in human cells (Xie et al. 2007). Nevertheless, CTCF binding motif is far from being fully characterized and additional information are continuously added that changes the canonical CTCF binding consensus sequence (Schmidt et al. 2012; Plasschaert et al. 2014). Very recently, additional consensus motives found upstream and downstream to the canonical CTCF-binding motif were reported (Nakahashi et al. 2014). The central motif is recognised by Zinc Fingers 4-7, while ZFs 9/11 bind to the upstream motif stabilizing the DNA-protein interaction.

The insulator property of the *Tyr* 5' boundary or LCR was previously described in our laboratory and was found associated with three sequences or motives, named A box, B box and G box (Ganss, Montoliu et al. 1994; Giraldo el al. 2003). Previous data shows that the sum of these three elements conveys the biological function, likely by cooperatively binding a series of transcription factors (Giraldo et al. 2003). Nevertheless, 3C data produced in this work indicate that only the *Dpn*II fragment containing the AB box is able to interact with the *Tyr*

promoter, and not the one containing the G box, suggesting that the role of this sequence might be less relevant that previously anticipated.

Furthermore, all mouse lines that we have generated, lacking the AB box and carrying the intact G box, display reduced *Tyr* expression. Further analysis of the chromatin marks *in vivo* will reveal if the loss of the *Tyr* 5' element results in the spreading of repressive chromatin marks from the LINE1 element till the *Tyr* promoter.

5.2 The *Tyr* locus in 3D

In this work, we described the structural architecture of the mouse *Tyr* locus using Chromosome Conformation Capture. In our model, two conserved elements, located at either ends of the *Tyr* coding sequence, contact the gene promoter (**Figure d1**). One of these interactions, involving the *Tyr* promoter and the *Tyr* 3' element, produces a constitutive DNA loop, found in both *Tyr* expressing and non-expressing cells, that is mediated by CTCF-binding. The interaction between *Tyr* promoter and the *Tyr* 5' boundary is cell-type specific, since it was detected more frequently in *Tyr* positive, B16 mouse melanoma cells. For this reason, we speculate that the melanocyte master transcription factor, *Mift* (which binds to three sites on the *Tyr* promoter on *Tyr*-expressing cells) interacts with other transcription factors bound to the 5' element, including AP-1, CREB (Giraldo et al. 2003) and CTCF (Angel Diaz PhD thesis, 2007). In addition, specific RNA molecules interacting with CTCF (as observed by Saldaña-Meyer et al. 2014), may be involved in the differential binding observed in the cell types studied.

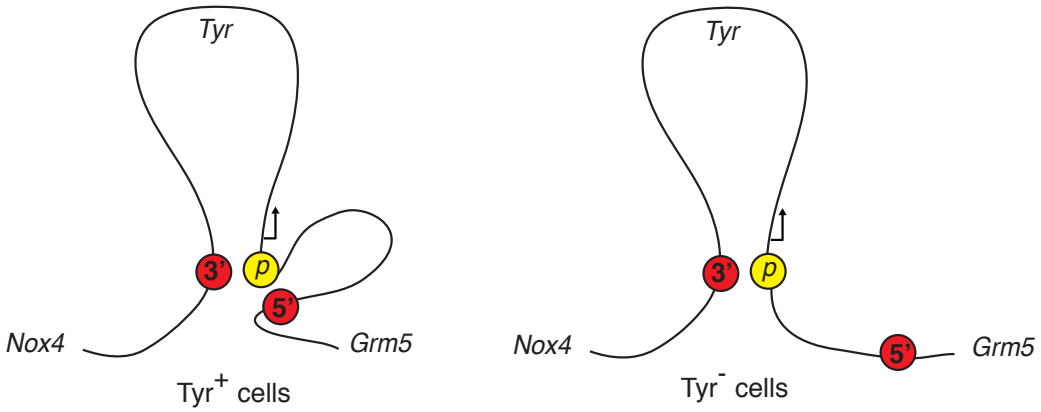


Figure d1 Model of the chromatin architecture of the mouse locus The locus is organized in one major loop constituted by the interaction between the *Tyr* promoter and the *Tyr* 3' element. In *Tyr*-expressing cells, a second loop is formed involving the *Tyr* 5' element.

Another interaction was found between the *Tyr* promoter and a sequence in the *Tyr* third intron, weaker but cell type-specific. The nature and the function of this interaction haven't been investigated in the present work, although a weak CTCF-binding could be observed in public ChIP-seq datasets.

Due to technical limitations, the architecture of the *Tyr* locus could not be described in the RPE cells, a second pigmented cell types, not of neural-crest origin. Both primary cultures and stable cell lines lack of *Tyr* expression. Our data, along with other works (Liu et al. 2009), remark that the transcriptional silencing of *Tyr* and of its master regulator, *Mitf*, is a consequence of the Epithelial-Mesenchymal transitions (EMT) that rapidly occurs upon culturing. In fact, we detect a strong expression of *Zeb1*, which is a transcriptional repressor associated with EMT, in both primary RPE cultures and in B6-RPE07 cells (**Figure r14**). The only chance to perform 3C on *Tyr*-expressing RPE cells would be to collect these cells directly from the mouse eye. Nevertheless, the procedure of eye dissection to access and collect these cells is time-consuming and yields a very small number of cells per eye. This discourages the *ex vivo* approach. One alternative, that was recently reported, would be to differentiate ES cells into RPE-like cells (Aruta et al. 2011; Bertolotti et al. 2014), a possibility that should be explored in future experiments.

The structure of other loci related with pigmentation has been studied recently. In humans, a key enhancer required for proper expression of OCA2 was found in the intron 80 of the flanking HERC2 gene, 21 kb upstream of OCA2. By 3C analysis, this element was found to be contacting the OCA2 promoter. SNPs associated with human skin pigmentation map on this sequence. In fact, darkly pigmented melanocytes (carrying the rs12913832 T-allele) show higher OCA2 expression and DNA looping between the enhancer and the OCA2 promoter, while lightly-pigmented melanocytes (carrying the C-allele), have reduced OCA2 expression and DNA looping (Visser et al 2012; Visser et al. 2014).
Similarly, other SNPs in non-coding region were found to be associated with hair pigmentation. In humans, Guenther and colleagues described a far-upstream KITLG (encoding KIT ligand) enhancer associated with blond hair colour (Guenther et al. 2014). In mice, chromosomal inversion of the orthologous region causes a lighter coat colour pigmentation (Bedell et al. 1995). All these examples underline the importance that non-coding DNA elements play in development and

disease and the fact that investigating gene loci from a structural point of view facilitates the discovery of relevant functional regulatory elements.

5.3 Why did TALENs fail?

TALENs represent the second generation of targeted nucleases after ZFNs. Rapidly after the first reports, the scientific community largely abandoned ZFNs to use the more versatile, cheaper and apparently more powerful TALENs (Chen et al. 2012). Nevertheless, in our experience, we could not observe a single *indel* event when we used TALENs against a chromosomal target, either *in vitro* or *in vivo*. On the contrary, we could observe nuclease activity when TALENs were used to digest purified PCR products (**Figure r15** and **r16**) in a cell-free assay or to target a multi-copy, episomal DNA template in transfected cells (**Figure r18** and **r19**). These results clearly indicate that (our) TALENs activities on single-copy chromosomal targets are limited. A reason could be the low affinity of TALENs with their substrate DNA sequence. In fact, we could rescue TALEN activity by providing a large excess of the substrate DNA. As an alternative, epigenetic marks present at the chromosomal target may be responsible of the lack of activity of our TALENs. In fact, these reagents appear to be unable to reach a target when the latter is found in repressive chromatin state (Ding et al. 2013; Hsu et al. 2013).

5.4 Why did CRISPR work?

After a series of test microinjections aimed to determine the optimal RNA concentration for embryo microinjection, we obtained with a number of founder animals larger than expected and larger than what we could handle and breed easily in our facility (22 potential founders out of 64 pups, **Table r10**). We hypothesize that the key of success using CRISPR depends on the way that the target recognition is mediated. On one hand, the *Watson&Crick* interaction between the guide RNA and the target DNA is in fact strong, predictable and energetically convenient. On the other hand, protein-DNA recognition, required using TALENs, is a less predictable and energy-consuming process. Although helicase activity of Cas9 was questioned (Mali et al. 2013), no energy-dependant activity was found. Possibly, the PAM binding could trigger a general

destabilization of the DNA duplex leading to the formation of the RNA-DNA heteroduplex (Sternberg et al. 2014).

5.5 Pay one, get more

When the error-prone non-homologous end-joining repair (NHEJ) pathway repairs a DSB, the resulting *DNA scar* varies according to micro-homologies found in the sequences flanking the DNA double-strand break (Lieber et al. 2003) and probably, according to which cell-cycle phase of the cell cycle is undergoing. As expected, when targeting two distal sites, in some lines we observed small sequence variations (in the order of tens of base pairs) around the break-point, where the two distal DNA ends are joined. What we did not expect to observe are the partial deletions, that we observed in both the TYRINS3 and TYRINS5 experiments, and the larger deletions, that spans hundreds of bases upstream to the targeted sequence. This variability in DNA repair has some advantages and, of course, some disadvantages. The main advantage, as suggested by the title of this paragraph, is that in one microinjection experiment (and using the same reagents) several distinct mouse lines can be generated. This is extremely useful for the study of non-coding regulatory element, where partial deletion can be used to restrict the DNA region that is relevant for a particular function. For example, we could take advantage of the line TYRINS5#8, carrying a partial, 0.7 kb deletion, to reinforce the concept that the AB boxes are relevant for the function of the *Tyr* 5' element. Similarly, larger deletion can be used to remark that the removal of additional DNA does not alter the phenotype observed with the exact or with the partial deletion. This is true for the line TYRINS5#13, that carries a larger, 1.7 kb deletion and is phenotypically indistinguishable from the other TYRINS5 lines. Moreover, chromosomal inversions can be generated that might be used to study if the function of the *Tyr* 5' element depends on its orientation. However, we could only confirm germline transmission in two out of seven founders.

All the additional mouse lines were generated using a number of animals that was calculated for the generation of a single line. Considering the 3Rs of animal experimentation ethics, this is a great example of Reduction. This can also be seen the other way around: the more efficient is genome editing in embryos, more transgenic lines will be generated, and more organism will be genetically altered and used in the laboratory (Combes and Balls, 2014).

On the other hand, the variability associated with CRISPR/Cas activity in embryos discourages its use in transient experiments, where it might be complicate to correlate an observed phenotype with a very unpredictable and complex genotype. In fact – considering our data - both the size of the deletion (**Figure r36**), the contribution of the edited cells to the organism, and the degree of mosaicism (**Figure r48** and **Figure d2**) can vary consistently between individuals obtained within the same microinjection session. With CRISPR, we and others (Yen et al. 2014) could observe a phenotype in founder animals, targeting recessive alleles. Nevertheless, founder animals generated with this technology can be genetically complex. We observed, for example, individuals carrying two different deletions (TYRINS5#18) or others carrying one deleted and one inverted allele (TYRINS5#30 and #60). All these genetic combinations can't be obviously used directly for experimentation but require accurate breeding.

5.6 Mosaicism

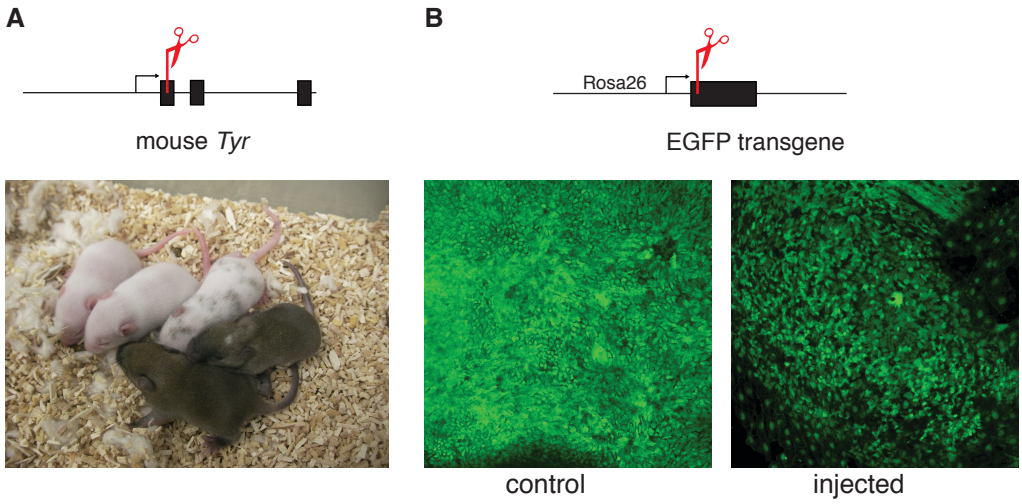


Figure d2 Mosaicism in CRISPR/Cas9-generated mice. **A** Litter obtained from the targeting of *Tyr* exon 1 in B6.CBA mouse embryos. Albino animals can be obtained (full, biallelic inactivation), as well as complex chimeras composed by wild-type cells, cells carrying a single targeting event and cells carrying both inactiveated copy of *Tyr*. **B** Targeting of a single copy EGFP transgene in embryos obtained from 129-Gt(ROSA)26Sor<tm1(CAG-EGFP)Luo/J mice. Confocal images of transgenic mouse fetuses. The number of $EGFP^{+}$ cells is reduced by injecting CRISPR but not fully abolished. These experiments were carried on in collaboration with Dr. Sagrario Ortega, CNIO.

In classical *additive* transgenesis, a mosaic is an organism in which the transgene integrated in the genome after the first cell divisions. As a result, the founder animal is composed of cells carrying the transgene and cells that do not harbour the transgene. In contrast, in nuclease-based genetic manipulation, a mosaic is an animal that carries multiple alleles, either a combination of a wild-

type and a mutated allele, or even a complex mixture of diverse mutated alleles, each of them having occurred in different stages and/or different cell lineages. In our experience with CRISPR, mosaicism is a frequent phenomenon. As shown in **Figure d2 A**, targeting *Tyr* coding sequence can lead to an albino phenotype (biallelic inactivation of the *Tyr* gene), or to patchy animals in which none, one or both copies of the gene have been targeted. Similarly, when we used CRISPR to target one single copy of EGFP (in heterozygous transgenic embryos), CRISPR activity after the first cell division resulted in the loss of GFP in many cells, but not all of them. Currently there are several strategies to reduce the degree of mosaicism: the use of Cas9 protein, and not Cas9 mRNA, could promote early cleavage of the target DNA (Gagnon et al. 2014). Interestingly, a Cas9-transgenic mouse has been recently reported (Platt et al. 2014). Injecting the sgRNA into oocytes that already contain Cas9 protein might increase transgenesis efficiency and reduce mosaicism.

5.7 Role of the *Tyr* 5' element *in vivo*

In this work we have proved the relevance of the *Tyr* 5' element in the endogenous locus. Similar but not conclusive evidences were collected using chromosome-type transgenes carrying deletion or mutation in the *Tyr* 5' element. In fact, *Tyr*-YAC lacking the full *Tyr* 5' element (YRT4 and ΔLCR, **Figure d3**) failed to fully rescue the pigmentation in albino background. These experiments were fundamentally useful to identify the presence of a key element (included in YAC YRT2 and YRT3, **Figure d3**), but technical limitations hampered the description of the core sequences in the *Tyr* 5' element. In fact, the inactivation of the A box and both A and B boxes in YACs (**Figure i7** and **Figure d3**), resulted in a variable reduction in pigmentation, not consistent in all the transgenic mouse lines generated using these constructs.

On the contrary, several independent lines generated with CRISPR targeting the *Tyr* 5' element showed robustly the same phenotype, as shown in **Figure d3** and **Figure r49**.

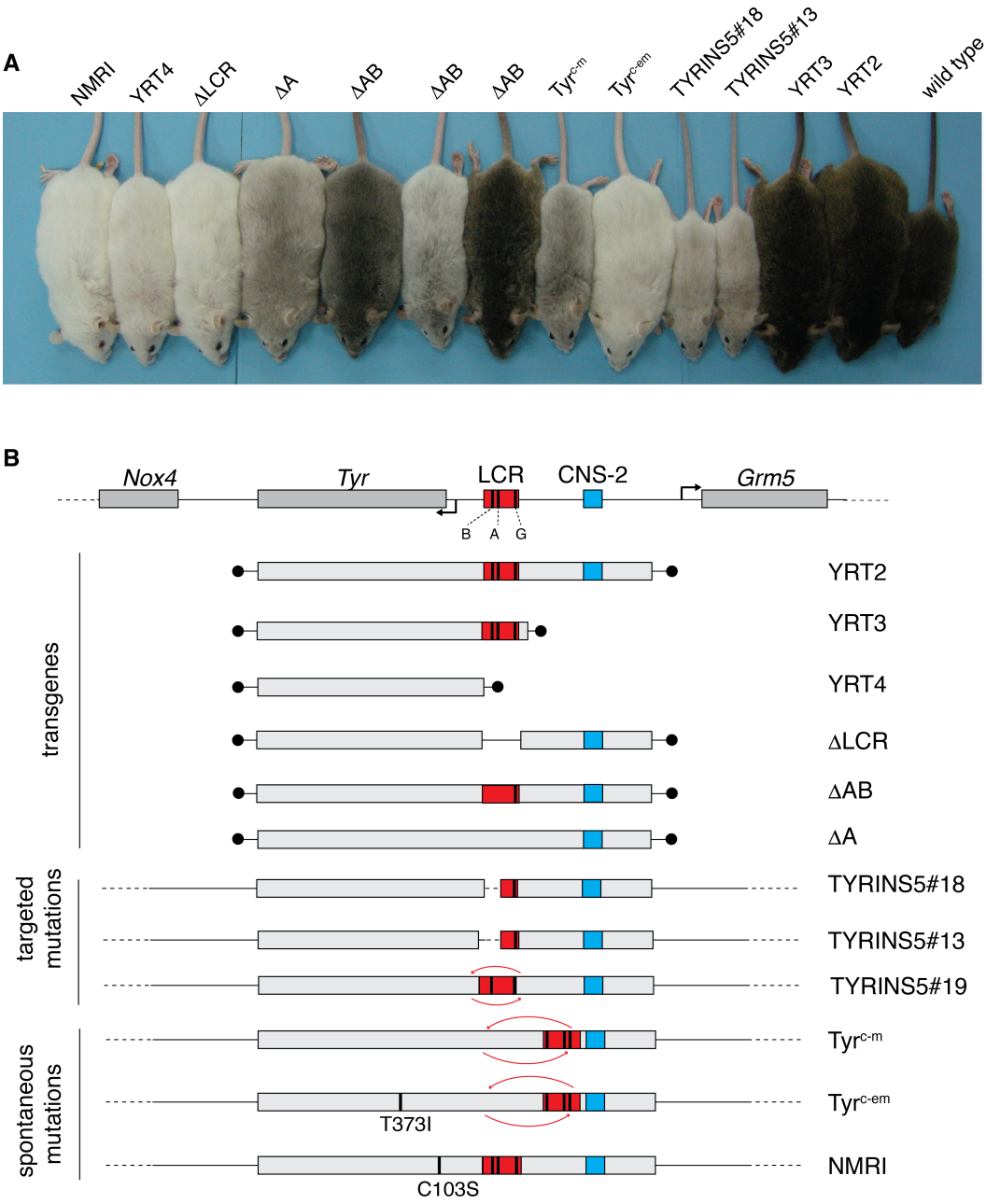


Figure d2 Comparison of the different *Tyr* transgenic mouse lines. **A** Individuals from the different mouse models generated to study the *Tyr* locus are shown. An albino NMRI mouse (left) and a B6CBAF2 mouse (right) are shown as albino and pigmented controls. Individuals from lines YRT4 and DLCR show pigmented eyes but lack of pigment in coat. A DA individual show a greyish coat colour. Individuals from three independent DAB lines show a variable phenotype. Lines TYRINS5#18 and #13 are indistinguishable. Spontaneous mutants *Tyr*$^{c\text{-}m}$ and *Tyr*$^{c\text{-}em}$ are also included, carrying a chromosomal rearrangement 5' to the *Tyr* coding sequence. Mice carrying the full YRT2 YAC and the YRT3 YAC (that includes the *Tyr* 5' element) show a pigmented phenotype indistinguishable from that of a wild-type pigmented individual. Animals are not age-matched. **B** Schematics of transgenes, targeted and spontaneous mutations of the mice shown in A.

Inactivating the *Tyr* 5' *element in vivo* produced a reduction of *Tyr* expression in skin and choroid melanocytes, two neural-crest derived cell lineages. On the contrary, we could not observe gross loss in pigmentation of the RPE layer of the retina, indicating that the *Tyr* 5' is dispensable to achieve *Tyr* expression in the RPE.

By comparing the phenotype of different lines carrying distinct deletion, we could identify a smaller sequence – within the *Tyr* 5' element – that is necessary for proper *Tyr* expression. The phenotype of lines #18a, #30, #50 and #60 is in fact indistinguishable from that of line #13, which carries a bigger deletion. We can thus state that the additional DNA missing in line #13 is not relevant for the *Tyr* 5' element. Similarly, comparing the lines carrying the full deletion with animals bearing a partial deletion, we could identify a small 410 bp that is always absent in all animals that show alteration in coat colour pigmentation (**Figure d4**).

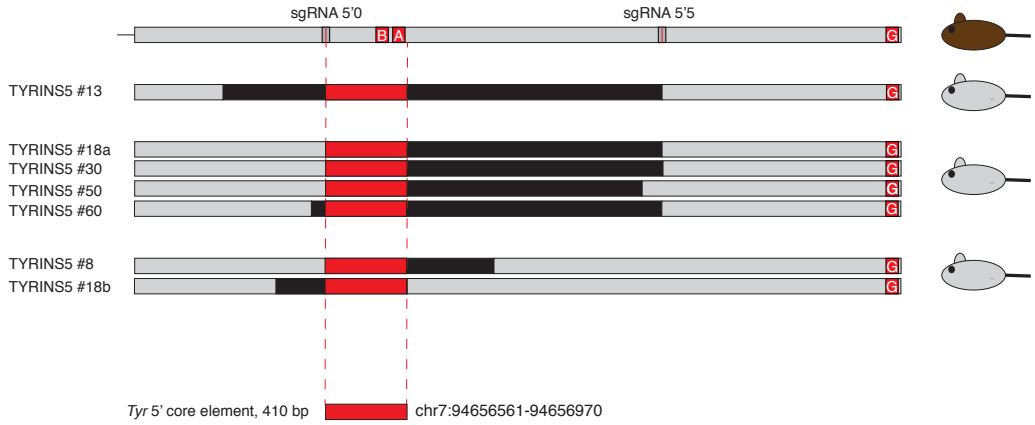


Figure d4 Identification of the *Tyr* 5' core sequence. By comparing the phenotype resulting from different deletions, a small, relevant, 410 bp large sequence was identified. It cannot be excluded that the size of this element can be further reduced without loosing the underlying phenotype. Genomic coordinates refer to the mm37 assembly.

Interestingly, the inversion of the Tyr 5' element caused the same phenotype observed upon its deletion. This suggests that not only the presence, but also the directionality of this element is fundamental. In fact, one possibility is that the Tyr 5' element has both enhancer and insulator properties. Upon inversion, the insulator element is positioned in between the promoter and the actual enhancer, blocking its activity. A similar inversion of the same DNA region was also observed in a spontaneous mouse mutant, the $Tyr^{c\text{-}m}$ (Porter et al. 1991) and in a related mutant $Tyr^{c\text{-}em}$ (Lavado et al. 2005), that shows alterations in coat colour similar to those observed in our *Tyr* 5' lines.

5.8 The role of the *Tyr* 3' element *in vivo*

The inactivation of one copy of the *Tyr* 3' element in albino background did not produced any evident alteration in coat colour, suggesting that this element is not necessary for the regulation of *Tyr*. One possibility is that this element, of which insulator activity has been detected in human cells and in zebrafish embryos, is required for the regulation of the flanking *Nox4* gene. This option is in contrast with our 3C data, in which the *Tyr* 3' element is clearly contacting the *Tyr* promoter. To verify this, mice carrying the *Tyr* 3' inactivation in both alleles must be analysed.

5.9 Reduced GLT of a chromosomal inversion

Targeting two distal DNA sites on the same chromosome led to the loss of the intervening DNA sequence. However, in a limited number of cases, we obtained the inversion of the DNA sequence flanked by the targeted sites. Similar results have been obtained *in vitro* by other groups (Choi et al. 2014). Furthermore, in our experiments, five out of seven founder animals carrying targeted inversions failed to transmit this genomic alteration to the F1 generation. We could observe germline transmission in just two founders out of seven (28%), TYRINS5#19 and #41, in contrast with the higher GTL observed for deletions (almost 100%). We therefore hypothesise that such inversion might cause problems in meiosis and, therefore, a reduced number functional gametes are produced carrying this edited allele (**Figure d5**).

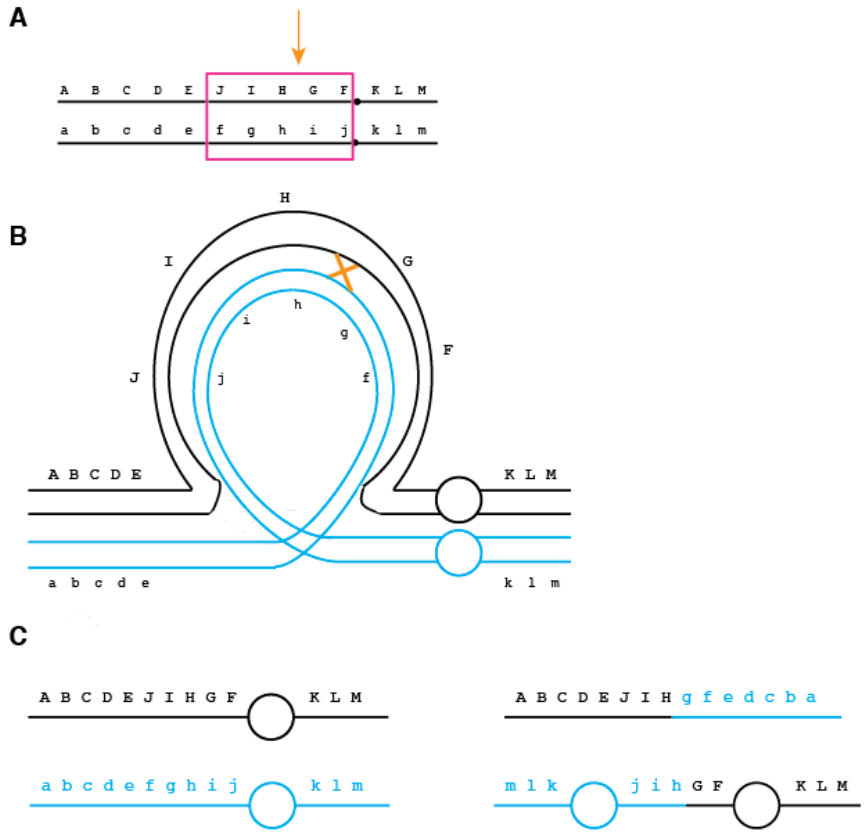


Figure d5 Gametes obtained by meiosis of a heterozygous inversion. **A** A locus contains an inversion indicated in the purple box. **B** If crossover happens at the position indicated by the orange cross, in correspondence with the inversion, the chromatids align forming a loop. **C** As a consequence, only half of the gametes produced are funtional. The other 50% is either acentric or dicentrinc.

Failure in germline transmission of the inversion was observed in both male and female founder mice. In humans, individuals carrying non-pathological inversion in autosomal chromosomes are described to be either infertile or subfertile (Rouen et al. 2013).

5.10 Experiments in progress and perspectives

The analysis of the TYRINS3 and TYRINS5 lines in albino background allowed us to detect the role of these elements with respect of *Tyr*. However, additional questions still remain to be answered. Does the *Tyr* 5' element influence the expression level and pattern of the flanking gene *Grm5*? Does the *Tyr* 3' element contribute to the regulatory elements of the *Nox4* gene? To answer these questions we are currently breeding our mouse lines to obtain homozygous animals lacking these regulatory elements. Should they be vital – we plan to analyse the expression of each of the three genes in this locus in distinct tissues. Also, further experiments are required to understand why the inversion of the *Tyr* 5' element result is a phenotype similar to that obtained upon complete deletion.

Does this element is directional? Do we interfere with *Grm5* upon inverting the *Tyr* 5' boundary? These and other questions will be addressed in future experiments.

6. Conclusions

6. CONCLUSIONS

1.- We identified a chromatin insulator 3' downstream the mouse *Tyr* gene using sequence conservation, synteny and published ChIP-seq datasets. We validated *in vitro* and *in vivo* the boundary activity of the *Tyr* 3' predicted insulator. This element is indeed a chromatin insulator whose activity mainly but not only relies on the binding of CTCF. We also confirmed *in vitro* the insulator activity of the *Tyr* 5' boundary, also known as *Tyr* LCR.

2.- The hub of the *Tyr* locus is the *Tyr* promoter. Both *Tyr* 3' and 5' boundaries, as well as other conserved elements found in the third *Tyr* intron, do contact the *Tyr* promoter.

3.- Our efforts to model the RPE cells *in vitro* were unsuccessful. The loss of RPE features, such as pigmentation, expression of specific markers and epithelial morphology associated with all established lines as well as with primary cultures prevented from using this approach for 3C experiments.

4.- Targeted nucleases can be used *in vivo* to obtain mutated alleles at non-coding sequences that were refractory to the modification using standard gene targeting in ES cells.

5.- By applying the CRISPR/Cas9 system (but not TALENs), we could obtain transgenic mouse lines carrying deletions of the *Tyr* 3' and *Tyr* 5' elements.

6.- Due to the error-prone NHEJ DNA repair pathway, we could obtain a series of transgenic mice carrying several different partial deletions of the *Tyr* 5' and 3'sequences of interest. This allowed us to narrow the minimal size of the regulatory elements in study. For the same reason, we also

obtained transgenic mice carrying the inversion of the *Tyr* 5' elements, allowing us to assess the directionality of this regulatory element.

7.- Mice lacking one copy of the *Tyr* 5' boundary in albino background element show a drastic reduction in coat colour pigmentation. At the histological level, we observed a reduced accumulation of melanin in the skin and variegation at the choroid in the eye as reported before for LCR-deleted YAC *Tyr*-transgenes. Pigmentation of the RPE layer does not appear affected. The *Tyr* 5' boundary seems to act as a lineage-specific, enhancer/insulator.

8.- Mice carrying an inversion of the *Tyr* 5' boundary in albino background have the same altered coat colour as those carrying a full deletion. Therefore, the directionality of this element in the genome appears to be essential to its function.

9.- Mice lacking one copy of *Tyr* 3' boundary in albino background do not show any alteration in eye or coat colour. This element appears to be therefore dispensable for regulation of *Tyr*. We could not exclude that such boundary is necessary for the proper expression of the flanking gene, *Nox4*.

7. CONCLUSIONES

1.- Hemos identificado una frontera genómica en la zona 3' del gen *Tyr* del genoma del ratón utilizando estrategias de conservación génica, sintenia y conjuntos de datos publicados de ChIP-seq. Hemos verificado in vitro e in vivo la actividad aisladora del elemento *Tyr* 3', postulado anteriormente como frontera genómica. Este elemento corresponde en realidad a una frontera genómica cuya actividad depende fundamentalmente de la unión con el factor CTCF. Hemos confirmado también la actividad in vitro del elemento aislador *Tyr* 5', también conocido como *Tyr* LCR.

2.- La zona principal del locus *Tyr* es su promotor. Las dos fronteras genómicas, *Tyr* 3' y *Tyr* 5', además de otros elementos reguladores conservados que se encuentran en el tercer intrón de *Tyr*, contactan con el promotor.

3.- No tuvimos éxito en nuestros esfuerzos para reproducir la función de las células RPE in vitro. La pérdida de las características como RPE, tales como la pigmentación, la expresión de marcadores específicos y su morfología epitelial, observadas en todas las líneas celulares establecidas, así como en cultivos primarios, nos impidió utilizar esta aproximación para los experimentos 3C.

4.- Las nucleasas de edición genómica pueden utilizarse in vivo para obtener alelos mutados en secuencias que no codifican para proteínas, y que normalmente se mostraban como recalcitrantes a su modificación mediante técnicas habituales de inactivación génica en células ES.

5.- Mediante la utilización del sistema CRISPR/Cas9 (pero no el sistema de TALENs) pudimos obtener líneas de ratones transgénicos portadoras de mutaciones en los elementos *Tyr* 3' y *Tyr* 5'.

6.- Gracias a la vía de reparación del ADN a través del sistema NHEJ, propenso a producir errores, pudimos obtener una serie de ratones transgénicos portadores de deleciones parciales de distinto tamaño tanto en el elemento *Tyr* 5' como en el elemento *Tyr* 3'. Esto nos ha permitido delimitar el tamaño mínimo de los elementos reguladores a estudiar. Por la misma razón, pudimos igualmente obtener ratones transgénicos portadores de inversiones del elemento *Tyr* 5', lo cual nos ha permitido investigar el efecto de la direccionabilidad de este elemento sobre su función reguladora.

7.- Los ratones que carecen de una copia del elemento aislador *Tyr* 5' en un fondo genético albino muestran una reducción significativa de la pigmentación del pelaje. A nivel histológico, observamos una reducción en la acumulación de melanina en la piel, y una variegación en el coroides del ojo, tal y como antes se había observado en los transgenes con YACs portadores de deleciones en la LCR. La pigmentación de las células RPE no parece estar afectada. El elemento aislador genómico *Tyr* 5' parece actuar de forma especifica de linaje, como un potenciador/aislador.

8.- Los ratones que portan una inversión del elemento *Tyr* 5' en un fondo genético albino muestran la misma alteración en la pigmentación que los que portan una deleción completa del mismo. Por lo tanto, la dirección de este elemento en el genoma parece ser esencial para su función.

9.- Los ratones que carecen de una copia del elemento *Tyr* 3', en un fondo genético albino, no muestran alteración pigmentaria alguna en el ojo o el pelaje. Este elemento no parece ser necesario para la regulación del gen *Tyr*. Sin embargo, no podemos excluir que dicho elemento aislador *Tyr* 3' sea necesario para la expression correcta del gen vecino Nox4.

8. References

1 Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, doi:10.1038/nature13579 (2014).
2 Andreasson, C. *et al.* Direct cloning of isogenic murine DNA in yeast and relevance of isogenicity for targeting in embryonic stem cells. *PloS one* **8**, e74207, doi:10.1371/journal.pone.0074207 (2013).
3 Arnheiter, H. The discovery of the microphthalmia locus and its gene, Mitf. *Pigment cell & melanoma research* **23**, 729-735, doi:10.1111/j.1755-148X.2010.00759.x (2010).
4 Aruta, C. *et al.* In vitro differentiation of retinal pigment epithelium from adult retinal stem cells. *Pigment cell & melanoma research* **24**, 233-240, doi:10.1111/j.1755-148X.2010.00793.x (2011).
5 Auer, T. O., Duroure, K., De Cian, A., Concordet, J. P. & Del Bene, F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome research* **24**, 142-153, doi:10.1101/gr.161638.113 (2014).
6 Baker, M. Repositories share key research tools. *Nature* **505**, 272, doi:10.1038/505272a (2014).
7 Barnes, D. E. Non-homologous end joining as a mechanism of DNA repair. *Current biology : CB* **11**, R455-457 (2001).
8 Bedell, M. A. *et al.* DNA rearrangements located over 100 kb 5' of the Steel (Sl)-coding region in Steel-panda and Steel-contrasted mice deregulate Sl expression and cause female sterility by disrupting ovarian follicle development. *Genes & development* **9**, 455-470 (1995).
9 Bedell, V. M. *et al.* In vivo genome editing using a high-efficiency TALEN system. *Nature* **491**, 114-118, doi:10.1038/nature11537 (2012).
10 Beerli, R. R. & Barbas, C. F., 3rd. Engineering polydactyl zinc-finger transcription factors. *Nature biotechnology* **20**, 135-141, doi:10.1038/nbt0202-135 (2002).
11 Beermann, F., Ruppert, S., Hummler, E. & Schutz, G. Tyrosinase as a marker for transgenic mice. *Nucleic acids research* **19**, 958 (1991).
12 Beermann, F., Schmid, E. & Schutz, G. Expression of the mouse tyrosinase gene during embryonic development: recapitulation of the temporal regulation in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 2809-2813 (1992).
13 Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387-396 (1999).
14 Bender, M. A., Bulger, M., Close, J. & Groudine, M. Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Molecular cell* **5**, 387-393 (2000).
15 Bentley, N. J., Eisen, T. & Goding, C. R. Melanocyte-specific expression of the human tyrosinase promoter: activation by the microphthalmia gene product and role of the initiator. *Molecular and cellular biology* **14**, 7996-

8006 (1994).
16 Bertolotti, E., Neri, A., Camparini, M., Macaluso, C. & Marigo, V. Stem cells as source for retinal pigment epithelium transplantation. *Progress in retinal and eye research* **42C**, 130-144, doi:10.1016/j.preteyeres.2014.06.002 (2014).
17 Bessa, J. *et al.* Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Developmental dynamics : an official publication of the American Association of Anatomists* **238**, 2409-2417, doi:10.1002/dvdy.22051 (2009).
18 Bhakta, M. S. *et al.* Highly active zinc-finger nucleases by extended modular assembly. *Genome research* **23**, 530-538, doi:10.1101/gr.143693.112 (2013).
19 Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* **161**, 1169-1175 (2002).
20 Boch, J. & Bonas, U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology* **48**, 419-436, doi:10.1146/annurev-phyto-080508-081936 (2010).
21 Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512, doi:10.1126/science.1178811 (2009).
22 Bolon, B. Whole mount enzyme histochemistry as a rapid screen at necropsy for expression of beta-galactosidase (LacZ)-bearing transgenes: considerations for separating specific LacZ activity from nonspecific (endogenous) galactosidase activity. *Toxicologic pathology* **36**, 265-276, doi:10.1177/0192623307312693 (2008).
23 Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561, doi:10.1099/mic.0.28048-0 (2005).
24 Bourque, G. *et al.* Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome research* **18**, 1752-1762, doi:10.1101/gr.080663.108 (2008).
25 Brito, F. C. & Kos, L. Timeline and distribution of melanocyte precursors in the mouse heart. *Pigment cell & melanoma research* **21**, 464-470, doi:10.1111/j.1755-148X.2008.00459.x (2008).
26 Buck, L. & Axel, R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell* **65**, 175-187 (1991).
27 Buhler, M., Verdel, A. & Moazed, D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873-886, doi:10.1016/j.cell.2006.04.025 (2006).
28 Bulger, M. *et al.* Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5129-5134 (1999).
29 Capecchi, M. R. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nature reviews. Genetics*

6, 507-512, doi:10.1038/nrg1619 (2005).
30 Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research* **39**, e82, doi:10.1093/nar/gkr218 (2011).
31 Chen, S. *et al.* A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic acids research* **41**, 2769-2778, doi:10.1093/nar/gks1356 (2013).
32 Cheng, G., Cao, Z., Xu, X., van Meir, E. G. & Lambeth, J. D. Homologs of gp91phox: cloning and tissue expression of Nox3, Nox4, and Nox5. *Gene* **269**, 131-140 (2001).
33 Choi, P. S. & Meyerson, M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nature communications* **5**, 3728, doi:10.1038/ncomms4728 (2014).
34 Chung, J. H., Whiteley, M. & Felsenfeld, G. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell* **74**, 505-514 (1993).
35 Chylinski, K., Le Rhun, A. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA biology* **10**, 726-737, doi:10.4161/rna.24321 (2013).
36 Combes, R. D. & Balls, M. Every silver lining has a cloud: the scientific and animal welfare issues surrounding a new approach to the production of transgenic animals. *Alternatives to laboratory animals : ATLA* **42**, 137-145 (2014).
37 Comet, I., Schuettengruber, B., Sexton, T. & Cavalli, G. A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2294-2299, doi:10.1073/pnas.1002059108 (2011).
38 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
39 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
40 Copeland, N. G., Jenkins, N. A. & Court, D. L. Recombineering: a powerful new tool for mouse functional genomics. *Nature reviews. Genetics* **2**, 769-779, doi:10.1038/35093556 (2001).
41 Danielian, P. S., Echelard, Y., Vassileva, G. & McMahon, A. P. A 5.5-kb enhancer is both necessary and sufficient for regulation of Wnt-1 transcription in vivo. *Developmental biology* **192**, 300-309, doi:10.1006/dbio.1997.8762 (1997).
42 Davis, A. A. *et al.* A human retinal pigment epithelial cell line that retains epithelial characteristics after prolonged culture. *Investigative ophthalmology & visual science* **36**, 955-964 (1995).
43 de la Calle-Mustienes, E. *et al.* A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. *Genome research* **15**, 1061-1072, doi:10.1101/gr.4004805 (2005).
44 de Laat, W. & Dekker, J. 3C-based technologies to study the shape of the

genome. *Methods* **58**, 189-191, doi:10.1016/j.ymeth.2012.11.005 (2012).
45 de Laat, W. & Grosveld, F. Spatial organization of gene expression: the active chromatin hub. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **11**, 447-459 (2003).
46 Dickins, R. A. *et al.* Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nature genetics* **37**, 1289-1295, doi:10.1038/ng1651 (2005).
47 Ding, Q. *et al.* Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell stem cell* **12**, 393-394, doi:10.1016/j.stem.2013.03.006 (2013).
48 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).
49 Doyle, E. L. *et al.* TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic acids research* **40**, W117-122, doi:10.1093/nar/gks608 (2012).
50 Doyon, Y. *et al.* Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nature methods* **8**, 74-79, doi:10.1038/nmeth.1539 (2011).
51 Drissen, R. *et al.* The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes & development* **18**, 2485-2490, doi:10.1101/gad.317004 (2004).
52 Dunn, K. C., Aotaki-Keen, A. E., Putkey, F. R. & Hjelmeland, L. M. ARPE-19, a human retinal pigment epithelial cell line with differentiated properties. *Experimental eye research* **62**, 155-169, doi:10.1006/exer.1996.0020 (1996).
53 Elizondo, L. I., Jafar-Nejad, P., Clewing, J. M. & Boerkoel, C. F. Gene clusters, molecular evolution and disease: a speculation. *Current genomics* **10**, 64-75, doi:10.2174/138920209787581271 (2009).
54 Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PloS one* **4**, e5553, doi:10.1371/journal.pone.0005553 (2009).
55 Ferguson, C. A. & Kidson, S. H. The regulation of tyrosinase gene transcription. *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society* **10**, 127-138 (1997).
56 Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811, doi:10.1038/35888 (1998).
57 Flemr, M. *et al.* A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. *Cell* **155**, 807-816, doi:10.1016/j.cell.2013.10.001 (2013).
58 Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826, doi:10.1038/nbt.2623 (2013).
59 Fujii, W., Kawasaki, K., Sugiura, K. & Naito, K. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease.

Nucleic acids research **41**, e187, doi:10.1093/nar/gkt772 (2013).
60 Gagnon, J. A. *et al.* Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PloS one* **9**, e98186, doi:10.1371/journal.pone.0098186 (2014).
61 Gaj, T., Gersbach, C. A. & Barbas, C. F., 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology* **31**, 397-405, doi:10.1016/j.tibtech.2013.04.004 (2013).
62 Ganss, R., Montoliu, L., Monaghan, A. P. & Schutz, G. A cell-specific enhancer far upstream of the mouse tyrosinase gene confers high level and copy number-related expression in transgenic mice. *The EMBO journal* **13**, 3083-3093 (1994).
63 Ganss, R., Schmidt, A., Schutz, G. & Beermann, F. Analysis of the mouse tyrosinase promoter in vitro and in vivo. *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society* **7**, 275-278 (1994).
64 Ganss, R., Schutz, G. & Beermann, F. The mouse tyrosinase gene. Promoter modulation by positive and negative regulatory elements. *The Journal of biological chemistry* **269**, 29808-29816 (1994).
65 Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71, doi:10.1038/nature09523 (2010).
66 Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2579-2586, doi:10.1073/pnas.1208507109 (2012).
67 Geurts, A. M. *et al.* Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325**, 433, doi:10.1126/science.1172447 (2009).
68 Gimenez, E., Giraldo, P., Jeffery, G. & Montoliu, L. Variegated expression and delayed retinal pigmentation during development in transgenic mice with a deletion in the locus control region of the tyrosinase gene. *Genesis* **30**, 21-25 (2001).
69 Gimenez, E., Lavado, A., Giraldo, P. & Montoliu, L. Tyrosinase gene expression is not detected in mouse brain outside the retinal pigment epithelium cells. *The European journal of neuroscience* **18**, 2673-2676 (2003).
70 Gimenez, E., Lavado, A., Jeffery, G. & Montoliu, L. Regional abnormalities in retinal development are associated with local ocular hypopigmentation. *The Journal of comparative neurology* **485**, 338-347, doi:10.1002/cne.20495 (2005).
71 Giraldo, P., Gimenez, E. & Montoliu, L. The use of yeast artificial chromosomes in transgenic animals: expression studies of the tyrosinase gene in transgenic mice. *Genetic analysis : biomolecular engineering* **15**, 175-178 (1999).
72 Giraldo, P. *et al.* Functional dissection of the mouse tyrosinase locus control region identifies a new putative boundary activity. *Nucleic acids research* **31**, 6290-6305 (2003).
73 Giraldo, P. & Montoliu, L. Size matters: use of YACs, BACs and PACs in

transgenic animals. *Transgenic research* **10**, 83-103 (2001).
74 Giraldo, P. & Montoliu, L. Artificial chromosome transgenesis in pigmentary research. *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society* **15**, 258-264 (2002).
75 Goding, C. R. Mitf from neural crest to melanoma: signal transduction and transcription in the melanocyte lineage. *Genes & development* **14**, 1712-1728 (2000).
76 Gonzalez, S. *et al.* Oncogenic activity of Cdc6 through repression of the INK4/ARF locus. *Nature* **440**, 702-706, doi:10.1038/nature04585 (2006).
77 Grosveld, F., Dillon, N. & Higgs, D. The regulation of human globin gene expression. *Bailliere's clinical haematology* **6**, 31-55 (1993).
78 Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**, 975-985 (1987).
79 Guenther, C. A., Tasic, B., Luo, L., Bedell, M. A. & Kingsley, D. M. A molecular basis for classic blond hair color in Europeans. *Nature genetics* **46**, 748-752, doi:10.1038/ng.2991 (2014).
80 Guo, J., Gaj, T. & Barbas, C. F., 3rd. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *Journal of molecular biology* **400**, 96-107, doi:10.1016/j.jmb.2010.04.060 (2010).
81 Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology* **1**, e60, doi:10.1371/journal.pcbi.0010060 (2005).
82 Hagege, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature protocols* **2**, 1722-1733, doi:10.1038/nprot.2007.243 (2007).
83 Halaban, R. *et al.* Tyrosinases of murine melanocytes with mutations at the albino locus. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 7241-7245 (1988).
84 Hall, C. V., Jacob, P. E., Ringold, G. M. & Lee, F. Expression and regulation of Escherichia coli lacZ gene fusions in mammalian cells. *Journal of molecular and applied genetics* **2**, 101-109 (1983).
85 Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355-1358, doi:10.1126/science.1192272 (2010).
86 Hermann, M., Cermak, T., Voytas, D. F. & Pelczar, P. Mouse genome engineering using designer nucleases. *Journal of visualized experiments : JoVE*, doi:10.3791/50930 (2014).
87 Hermann, M. *et al.* Evaluation of OPEN zinc finger nucleases for direct gene targeting of the ROSA26 locus in mouse embryos. *PloS one* **7**, e41796, doi:10.1371/journal.pone.0041796 (2012).
88 Hodgkinson, C. A. *et al.* Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. *Cell* **74**, 395-404 (1993).

89 Horii, T. *et al.* Validation of microinjection methods for generating knockout mice by CRISPR/Cas-mediated genome engineering. *Scientific reports* **4**, 4513, doi:10.1038/srep04513 (2014).
90 Horii, T. *et al.* Genome engineering of mammalian haploid embryonic stem cells using the Cas9/RNA system. *PeerJ* **1**, e230, doi:10.7717/peerj.230 (2013).
91 Horvath, L. M., Li, N. & Carrel, L. Deletion of an X-inactivation boundary disrupts adjacent gene silencing. *PLoS genetics* **9**, e1003952, doi:10.1371/journal.pgen.1003952 (2013).
92 Hou, C., Zhao, H., Tanimoto, K. & Dean, A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20398-20403, doi:10.1073/pnas.0808506106 (2008).
93 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832, doi:10.1038/nbt.2647 (2013).
94 Jackson, I. J. & Bennett, D. C. Identification of the albino mutation of mouse tyrosinase by analysis of an in vitro revertant. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 7010-7014 (1990).
95 Jansen, R., Embden, J. D., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* **43**, 1565-1575 (2002).
96 Jao, L. E., Wente, S. R. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 13904-13909, doi:10.1073/pnas.1308335110 (2013).
97 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
98 Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471, doi:10.7554/eLife.00471 (2013).
99 Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997, doi:10.1126/science.1247997 (2014).
100 Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nature reviews. Molecular cell biology* **14**, 49-55, doi:10.1038/nrm3486 (2013).
101 Kawakami, K. Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element. *Methods in cell biology* **77**, 201-222 (2004).
102 Kawasaki, H., Taira, K. & Morris, K. V. siRNA induced transcriptional gene silencing in mammalian cells. *Cell cycle* **4**, 442-448 (2005).
103 Kellum, R. & Schedl, P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**, 941-950 (1991).
104 Kettleborough, R. N., Bruijn, E., Eeden, F., Cuppen, E. & Stemple, D. L. High-throughput target-selected gene inactivation in zebrafish. *Methods in cell biology* **104**, 121-127, doi:10.1016/B978-0-12-374814-0.00006-9 (2011).

105 Kim, D. H., Villeneuve, L. M., Morris, K. V. & Rossi, J. J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nature structural & molecular biology* **13**, 793-797, doi:10.1038/nsmb1142 (2006).
106 King, R. A. *et al.* Tyrosinase gene mutations in oculocutaneous albinism 1 (OCA1): definition of the phenotype. *Human genetics* **113**, 502-513, doi:10.1007/s00439-003-0998-1 (2003).
107 Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10171-10176, doi:10.1073/pnas.0900604106 (2009).
108 Kress, C., Vandormael-Pournin, S., Baldacci, P., Cohen-Tannoudji, M. & Babinet, C. Nonpermissiveness for mouse embryonic stem (ES) cell derivation circumvented by a single backcross to 129/Sv strain: establishment of ES cell lines bearing the Omd conditional lethal mutation. *Mammalian genome : official journal of the International Mammalian Genome Society* **9**, 998-1001 (1998).
109 Kwon, B. S. *et al.* Isolation, chromosomal mapping, and expression of the mouse tyrosinase gene. *The Journal of investigative dermatology* **93**, 589-594 (1989).
110 Kwon, B. S., Wakulchik, M., Haq, A. K., Halaban, R. & Kestler, D. Sequence analysis of mouse tyrosinase cDNA and the effect of melanotropin on its gene expression. *Biochemical and biophysical research communications* **153**, 1301-1309 (1988).
111 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).
112 Lay, J. M., Friis-Hansen, L., Gillespie, P. J. & Samuelson, L. C. Rapid confirmation of gene targeting in embryonic stem cells using two long-range PCR techniques. *Transgenic research* **7**, 135-140 (1998).
113 Li, D. *et al.* Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nature biotechnology* **31**, 681-683, doi:10.1038/nbt.2661 (2013).
114 Li, W. *et al.* Genetic modification and screening in rat using haploid embryonic stem cells. *Cell stem cell* **14**, 404-414, doi:10.1016/j.stem.2013.11.016 (2014).
115 Li, W. *et al.* Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature* **490**, 407-411, doi:10.1038/nature11435 (2012).
116 Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nature reviews. Molecular cell biology* **4**, 712-720, doi:10.1038/nrm1202 (2003).
117 Lim, S. *et al.* A simple strategy for heritable chromosomal deletions in zebrafish via the combinatorial action of targeting nucleases. *Genome biology* **14**, R69, doi:10.1186/gb-2013-14-7-r69 (2013).
118 Liu, P. Q. *et al.* Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnology and bioengineering* **106**, 97-105, doi:10.1002/bit.22654 (2010).

119 Liu, Q., Segal, D. J., Ghiara, J. B. & Barbas, C. F., 3rd. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5525-5530 (1997).
120 Liu, Y. *et al.* Inheritable and precise large genomic deletions of non-coding RNA genes in zebrafish using TALENs. *PloS one* **8**, e76387, doi:10.1371/journal.pone.0076387 (2013).
121 Liu, Y. *et al.* Zeb1 represses Mitf and regulates pigment synthesis, cell proliferation, and epithelial morphology. *Investigative ophthalmology & visual science* **50**, 5080-5088, doi:10.1167/iovs.08-2911 (2009).
122 Lowings, P., Yavuzer, U. & Goding, C. R. Positive and negative elements regulate a melanocyte-specific promoter. *Molecular and cellular biology* **12**, 3653-3662 (1992).
123 Lunyak, V. V. Boundaries. Boundaries...Boundaries??? *Current opinion in cell biology* **20**, 281-287, doi:10.1016/j.ceb.2008.03.018 (2008).
124 Lunyak, V. V. *et al.* Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317**, 248-251, doi:10.1126/science.1140871 (2007).
125 Ma, A. C., Lee, H. B., Clark, K. J. & Ekker, S. C. High efficiency In Vivo genome engineering with a simplified 15-RVD GoldyTALEN design. *PloS one* **8**, e65259, doi:10.1371/journal.pone.0065259 (2013).
126 Ma, S. *et al.* Multiplex genomic structure variation mediated by TALEN and ssODN. *BMC genomics* **15**, 41, doi:10.1186/1471-2164-15-41 (2014).
127 Ma, Y. *et al.* Heritable multiplex genetic engineering in rats using CRISPR/Cas9. *PloS one* **9**, e89413, doi:10.1371/journal.pone.0089413 (2014).
128 MacGregor, G. R. & Caskey, C. T. Construction of plasmids that express E. coli beta-galactosidase in mammalian cells. *Nucleic acids research* **17**, 2365 (1989).
129 Maeder, M. L. *et al.* Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell* **31**, 294-301, doi:10.1016/j.molcel.2008.06.016 (2008).
130 Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology* **9**, 467-477, doi:10.1038/nrmicro2577 (2011).
131 Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* **31**, 833-838, doi:10.1038/nbt.2675 (2013).
132 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).
133 Maresca, M., Lin, V. G., Guo, N. & Yang, Y. Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome research* **23**, 539-546, doi:10.1101/gr.145441.112 (2013).
134 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845, doi:10.1126/science.1165771 (2008).
135 Martin, D. *et al.* Genome-wide CTCF distribution in vertebrates defines

equivalent sites that aid the identification of disease-associated genes. *Nature structural & molecular biology* **18**, 708-714, doi:10.1038/nsmb.2059 (2011).
136 Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D. & Wolfe, S. A. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nature biotechnology* **26**, 695-701, doi:10.1038/nbt1398 (2008).
137 Mohun, T. J., Garrett, N. & Gurdon, J. B. Upstream sequences required for tissue-specific activation of the cardiac actin gene in Xenopus laevis embryos. *The EMBO journal* **5**, 3185-3193 (1986).
138 Molto, E., Fernandez, A. & Montoliu, L. Boundaries in vertebrate genomes: different solutions to adequately insulate gene expression domains. *Briefings in functional genomics & proteomics* **8**, 283-296, doi:10.1093/bfgp/elp031 (2009).
139 Montoliu, L., Chavez, S. & Vidal, M. Variegation associated with lacZ in transgenic animals: a warning note. *Transgenic research* **9**, 237-239 (2000).
140 Montoliu, L. *et al.* Increasing the complexity: new genes and new types of albinism. *Pigment cell & melanoma research* **27**, 11-18, doi:10.1111/pcmr.12167 (2014).
141 Montoliu, L., Larue, L. & Beermann, F. On the use of regulatory regions from pigmentary genes to drive the expression of transgenes in mice. *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society* **17**, 188-190, doi:10.1046/j.1600-0749.2003.00124.x (2004).
142 Montoliu, L., Umland, T. & Schutz, G. A locus control region at -12 kb of the tyrosinase gene. *The EMBO journal* **15**, 6026-6034 (1996).
143 Moore, M., Choo, Y. & Klug, A. Design of polyzinc finger peptides with structured linkers. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 1432-1436, doi:10.1073/pnas.98.4.1432 (2001).
144 Moreira, P. N. *et al.* Efficient generation of transgenic mice with intact yeast artificial chromosomes by intracytoplasmic sperm injection. *Biology of reproduction* **71**, 1943-1947, doi:10.1095/biolreprod.104.032904 (2004).
145 Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289-1292, doi:10.1126/science.1101372 (2004).
146 Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501, doi:10.1126/science.1178817 (2009).
147 Muller, G., Ruppert, S., Schmid, E. & Schutz, G. Functional analysis of alternatively spliced tyrosinase gene transcripts. *The EMBO journal* **7**, 2723-2730 (1988).
148 Murisier, F., Guichard, S. & Beermann, F. Distinct distal regulatory elements control tyrosinase expression in melanocytes and the retinal pigment epithelium. *Developmental biology* **303**, 838-847, doi:10.1016/j.ydbio.2006.11.038 (2007).
149 Nagy, A. Cre recombinase: the universal reagent for genome tailoring. *Genesis* **26**, 99-109 (2000).

150 Nakahashi, H. *et al.* A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell reports* **3**, 1678-1689, doi:10.1016/j.celrep.2013.04.024 (2013).
151 Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949, doi:10.1016/j.cell.2014.02.001 (2014).
152 Olmsted, J. B., Carlson, K., Klebe, R., Ruddle, F. & Rosenbaum, J. Isolation of microtubule protein from cultured mouse neuroblastoma cells. *Proceedings of the National Academy of Sciences of the United States of America* **65**, 129-136 (1970).
153 Osoegawa, K. *et al.* Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome research* **10**, 116-128 (2000).
154 Palstra, R. J. *et al.* The beta-globin nuclear compartment in development and erythroid differentiation. *Nature genetics* **35**, 190-194, doi:10.1038/ng1244 (2003).
155 Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422-433, doi:10.1016/j.cell.2008.01.011 (2008).
156 Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology* **31**, 839-843, doi:10.1038/nbt.2673 (2013).
157 Peterson, K. R. *et al.* Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6605-6609 (1996).
158 Petolino, J. F. *et al.* Zinc finger nuclease-mediated transgene deletion. *Plant molecular biology* **73**, 617-628, doi:10.1007/s11103-010-9641-4 (2010).
159 Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194-1211, doi:10.1016/j.cell.2009.06.001 (2009).
160 Plasschaert, R. N. *et al.* CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. *Nucleic acids research* **42**, 774-789, doi:10.1093/nar/gkt910 (2014).
161 Platt, R. J. *et al.* CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. *Cell* **159**, 440-455, doi:10.1016/j.cell.2014.09.014 (2014).
162 Porter, S., Larue, L. & Mintz, B. Mosaicism of tyrosinase-locus transcription and chromatin structure in dark vs. light melanocyte clones of homozygous chinchilla-mottled mice. *Developmental genetics* **12**, 393-402, doi:10.1002/dvg.1020120604 (1991).
163 Porter, S. D. & Meyer, C. J. A distal tyrosinase upstream element stimulates gene expression in neural-crest-derived melanocytes of transgenic mice: position-independent and mosaic expression. *Development* **120**, 2103-2111 (1994).
164 Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389,

doi:10.1016/j.cell.2013.08.021 (2013).
165 Recillas-Targa, F., Bell, A. C. & Felsenfeld, G. Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 14354-14359 (1999).
166 Recillas-Targa, F. *et al.* Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6883-6888, doi:10.1073/pnas.102179399 (2002).
167 Reeves, R. H., Pavan, W. J. & Hieter, P. Yeast artificial chromosome modification and manipulation. *Methods in enzymology* **216**, 584-603 (1992).
168 Reid, L. H., Gregg, R. G., Smithies, O. & Koller, B. H. Regulatory elements in the introns of the human HPRT gene are necessary for its expression in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 4299-4303 (1990).
169 Reid, L. H., Shesely, E. G., Kim, H. S. & Smithies, O. Cotransformation and gene targeting in mouse embryonic stem cells. *Molecular and cellular biology* **11**, 2769-2777 (1991).
170 Roman, A. C. *et al.* Dioxin receptor and SLUG transcription factors regulate the insulator activity of B1 SINE retrotransposons via an RNA polymerase switch. *Genome research* **21**, 422-432, doi:10.1101/gr.111203.110 (2011).
171 Rosenberg, W. S., Breakefield, X. O., DeAntonio, C. & Isacson, O. Authentic and artifactual detection of the E. coli lacZ gene product in the rat brain by histochemical methods. *Brain research. Molecular brain research* **16**, 311-315 (1992).
172 Rouen, A. *et al.* Simultaneous cell by cell study of both DNA fragmentation and chromosomal segregation in spermatozoa from chromosomal rearrangement carriers. *Journal of assisted reproduction and genetics* **30**, 383-390, doi:10.1007/s10815-012-9915-7 (2013).
173 Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287**, 2204-2215 (2000).
174 Ruppert, S., Muller, G., Kwon, B. & Schutz, G. Multiple transcripts of the mouse tyrosinase gene are generated by alternative splicing. *The EMBO journal* **7**, 2715-2722 (1988).
175 Ryder, E. *et al.* Rapid conversion of EUCOMM/KOMP-CSD alleles in mouse embryos using a cell-permeable Cre recombinase. *Transgenic research* **23**, 177-185, doi:10.1007/s11248-013-9764-x (2014).
176 Saldana-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes & development* **28**, 723-734, doi:10.1101/gad.236869.113 (2014).
177 Schedl, A. *et al.* Transgenic mice generated by pronuclear injection of a yeast artificial chromosome. *Nucleic acids research* **20**, 3073-3077 (1992).
178 Schedl, A., Montoliu, L., Kelsey, G. & Schutz, G. A yeast artificial chromosome covering the tyrosinase gene confers copy number-dependent expression in transgenic mice. *Nature* **362**, 258-261, doi:10.1038/362258a0 (1993).

179 Schmidt, D. *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335-348, doi:10.1016/j.cell.2011.11.058 (2012).
180 Seruggia, D. & Montoliu, L. The new CRISPR-Cas system: RNA-guided genome engineering to efficiently produce any desired genetic alteration in animals. *Transgenic research*, doi:10.1007/s11248-014-9823-y (2014).
181 Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458-472, doi:10.1016/j.cell.2012.01.010 (2012).
182 Shaw, G., Morse, S., Ararat, M. & Graham, F. L. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **16**, 869-871, doi:10.1096/fj.01-0995fje (2002).
183 Shen, B. *et al.* Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell research* **23**, 720-723, doi:10.1038/cr.2013.46 (2013).
184 Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature methods* **11**, 399-402, doi:10.1038/nmeth.2857 (2014).
185 Shibata, K., Muraosa, Y., Tomita, Y., Tagami, H. & Shibahara, S. Identification of a cis-acting element that enhances the pigment cell-specific expression of the human tyrosinase gene. *The Journal of biological chemistry* **267**, 20584-20588 (1992).
186 Shim, S., Kwan, K. Y., Li, M., Lefebvre, V. & Sestan, N. Cis-regulatory control of corticospinal system development and evolution. *Nature* **486**, 74-79, doi:10.1038/nature11094 (2012).
187 Sollu, C. *et al.* Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic acids research* **38**, 8269-8276, doi:10.1093/nar/gkq720 (2010).
188 Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67, doi:10.1038/nature13011 (2014).
189 Stief, A., Winter, D. M., Stratling, W. H. & Sippel, A. E. A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature* **341**, 343-345, doi:10.1038/341343a0 (1989).
190 Tanaka, S., Yamamoto, H., Takeuchi, S. & Takeuchi, T. Melanization in albino mice transformed by introducing cloned mouse tyrosinase gene. *Development* **108**, 223-227 (1990).
191 te Riele, H., Maandag, E. R. & Berns, A. Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 5128-5132 (1992).
192 Tempel, S. Using and understanding RepeatMasker. *Methods in molecular biology* **859**, 29-51, doi:10.1007/978-1-61779-603-6_2 (2012).
193 Tiana, M. *et al.* A role for insulator elements in the regulation of gene expression response to hypoxia. *Nucleic acids research* **40**, 1916-1927, doi:10.1093/nar/gkr842 (2012).
194 Tief, K., Schmidt, A., Aguzzi, A. & Beermann, F. Tyrosinase is a new

marker for cell populations in the mouse neural tube. *Developmental dynamics : an official publication of the American Association of Anatomists* **205**, 445-456, doi:10.1002/(SICI)1097-0177(199604)205:4<445::AID-AJA8>3.0.CO;2-I (1996).

195 Torres, R. *et al.* Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nature communications* **5**, 3964, doi:10.1038/ncomms4964 (2014).

196 Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics* **11**, 636-646, doi:10.1038/nrg2842 (2010).

197 van den Boogaard, M. *et al.* A common genetic variant within SCN10A modulates cardiac SCN5A expression. *The Journal of clinical investigation* **124**, 1844-1852, doi:10.1172/JCI73140 (2014).

198 Van Keuren, M. L., Gavrilina, G. B., Filipiak, W. E., Zeidler, M. G. & Saunders, T. L. Generating transgenic mice from bacterial artificial chromosomes: transgenesis efficiency, integration and expression outcomes. *Transgenic research* **18**, 769-785, doi:10.1007/s11248-009-9271-2 (2009).

199 Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672-676, doi:10.1126/science.1093686 (2004).

200 Visel, A. *et al.* Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature* **464**, 409-412, doi:10.1038/nature08801 (2010).

201 Visser, M., Kayser, M., Grosveld, F. & Palstra, R. J. Genetic variation in regulatory DNA elements: the case of OCA2 transcriptional regulation. *Pigment cell & melanoma research* **27**, 169-177, doi:10.1111/pcmr.12210 (2014).

202 Visser, M., Kayser, M. & Palstra, R. J. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome research* **22**, 446-455, doi:10.1101/gr.128652.111 (2012).

203 Volpi, S. A. *et al.* Germline deletion of Igh 3' regulatory region elements hs 5, 6, 7 (hs5-7) affects B cell-specific regulation, rearrangement, and insulation of the Igh locus. *Journal of immunology* **188**, 2556-2566, doi:10.4049/jimmunol.1102763 (2012).

204 Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910-918, doi:10.1016/j.cell.2013.04.025 (2013).

205 Watanabe, S., Watanabe, S., Sakamoto, N., Sato, M. & Akasaka, K. Functional analysis of the sea urchin-derived arylsulfatase (Ars)-element in mammalian cells. *Genes to cells : devoted to molecular & cellular mechanisms* **11**, 1009-1021, doi:10.1111/j.1365-2443.2006.00996.x (2006).

206 Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796-801, doi:10.1038/nature06634 (2008).

207 West, A. G., Gaszner, M. & Felsenfeld, G. Insulators: many functions, many mechanisms. *Genes & development* **16**, 271-288, doi:10.1101/gad.954702 (2002).

208 Williams, A. & Flavell, R. A. The role of CTCF in regulating nuclear organization. *The Journal of experimental medicine* **205**, 747-750, doi:10.1084/jem.20080066 (2008).
209 Wilson, C., Bellen, H. J. & Gehring, W. J. Position effects on eukaryotic gene expression. *Annual review of cell biology* **6**, 679-714, doi:10.1146/annurev.cb.06.110190.003335 (1990).
210 Winder, D. G. & Conn, P. J. Roles of metabotropic glutamate receptors in glial function and glial-neuronal communication. *Journal of neuroscience research* **46**, 131-137, doi:10.1002/(SICI)1097-4547(19961015)46:2<131::AID-JNR1>3.0.CO;2-I (1996).
211 Xiao, A. *et al.* Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic acids research* **41**, e141, doi:10.1093/nar/gkt464 (2013).
212 Xie, X. *et al.* Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 7145-7150, doi:10.1073/pnas.0701811104 (2007).
213 Yamamoto, H., Takeuchi, S., Kudo, T., Sato, C. & Takeuchi, T. Melanin production in cultured albino melanocytes transfected with mouse tyrosinase cDNA. *Idengaku zasshi* **64**, 121-135 (1989).
214 Yanagisawa, H., Clouthier, D. E., Richardson, J. A., Charite, J. & Olson, E. N. Targeted deletion of a branchial arch-specific enhancer reveals a role of dHAND in craniofacial development. *Development* **130**, 1069-1078 (2003).
215 Yang, H., Wang, H. & Jaenisch, R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nature protocols* **9**, 1956-1968, doi:10.1038/nprot.2014.134 (2014).
216 Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370-1379, doi:10.1016/j.cell.2013.08.022 (2013).
217 Yang, J. & Corces, V. G. Insulators, long-range interactions, and genome function. *Current opinion in genetics & development* **22**, 86-92, doi:10.1016/j.gde.2011.12.007 (2012).
218 Yasumoto, K., Yokoyama, K., Shibata, K., Tomita, Y. & Shibahara, S. Microphthalmia-associated transcription factor as a regulator for melanocyte-specific transcription of the human tyrosinase gene. *Molecular and cellular biology* **14**, 8058-8070 (1994).
219 Yen, S. T. *et al.* Somatic mosaicism and allele complexity induced by CRISPR/Cas9 RNA injections in mouse zygotes. *Developmental biology* **393**, 3-9, doi:10.1016/j.ydbio.2014.06.017 (2014).
220 Yokoyama, T. *et al.* Conserved cysteine to serine mutation in tyrosinase is responsible for the classical albino mutation in laboratory mice. *Nucleic acids research* **18**, 7293-7298 (1990).
221 Zhou, J. *et al.* One-step generation of different immunodeficient mice with multiple gene modifications by CRISPR/Cas9 mediated genome engineering. *The international journal of biochemistry & cell biology* **46**, 49-55, doi:10.1016/j.biocel.2013.10.010 (2014).
222 Maniatis, T *et al.* Molecular cloning : a laboratory manual. *Cold Spring Harbor*, N.Y. : Cold Spring Harbor Laboratory Press (1982).

223 Montoliu, L Generation of Transgenic Mice (1997). cnb.csic.es/~montoliu/transgenic.html
224 Behringer, R *et al.* Manipulating the mouse embryo – A laboratory manual. *Cold Spring Harbor*, N.Y. : Cold Spring Harbor Laboratory Press (2014).

9. Appendices

9.1: Oligonucleotides used in this study

9.1.1 Oligonucleotides used to clone the mouse *Tyr* 3' element into the pELuc vector

XhoI_Tyr3'Fw	CAGCTCGAGACAGAAATGGCCCCACCTAT
XhoI_Tyr3'Rv	CAGCTCGAGTGCATTTGAACTTGACCTACTGA
PstI_Tyr3'Fw	CAGCTGCAGACAGAAATGGCCCCACCTAT
PstI_Tyr3'Rv	CAGCTGCAGTGCATTTGAACTTGACCTACTGA
XhoI_Tyr3'sineb2Fw	CAGCTCGAGCCAGGTGAGGGGTGTGTTTA
XhoI_Tyr3'sineb2Rv	CAGCTCGAGGAAGTGTTTATTGACAATGTG
PstI_Tyr3'sineb2Fw	CAGCTGCAGCCAGGTGAGGGGTGTGTTTA
PstI_Tyr3'sinb2Rv	CAGCTGCAGGAAGTGTTTATTGACAATGTG

9.1.2 Oligonucleotides used for site-directed mutagenesis of the *Tyr* 3' CTCF binding site

3'CTCFmutFw	TGTCTTCAGACACTCGAGAATAGAGCGCCAGATCTTGTTA
3'CTCFmutRv	TAACAAGATCTGGCGCTCTATTCTCGAGTGTCTGAAGACA

9.1.3 Oligonucleotides used for mouse genotyping, *in vitro* cleavage assay and T7 Endonuclease I assay

delLCRFW	CAACCAGGCTTTCATCAGAAT
delLCRRV	TTTTCTCTGTATCATGATTGCCTA
del3'FW	TCTGTGCATGGTATACAACAGGG
del3'RV	TGTACTCATGACCCCATTTTGACA
T7EILCR123FW	AACAGCTTCCCTAGAAGCTCAA
T7EILCR123RV	AGAACCAGCTTGAGTGTGTCA
T7EILCR4FW	AATATTGTGGTTTGCCAGGACC
T7EILCR4RV	TCGTGGGTCCACAAATCCATTG
T7EI3'56FW	TCTGCTGTGTTCCAATTTGCAG
T7EI3'56RV	GTGCATTAAAGGAAGCCCAATGA
T7EI3'123RV	TGGAACACAGTGTATTAAAAAGCCA
LCR67	TGAGCCTAAGAGACAAGATTG
LCRprobeFw	AGCTGTGCAGCCATTTTCCAAG
5'0OT1_Fw	AGACCATTGCTGACCTCTTACAG
5'0OT1_Rv	ACTTTGCTCAGCCCCACATAATA
5'0OT2_Fw	AAGGTACAGATAGAAACGAGCAGTT
5'0OT2_Rv	AGTGGCTAAGGAAACAAAGAATGTG
5'0OT3_Fw	GGGTTCAGACAGGTCAACTATGT
5'0OT3_Rv	GTGAAGTGCCATAAGCTGTGTTT
5'5OT1_Fw	CAGGTTAGAGGCCTAGCCAATAG
5'5OT1_Rv	GGAATACTGAGGACAGGTGACTG
5'5OT2_Fw	AGGAATATGACGGCAGCATTAGT
5'5OT2_Rv	GAGGAAGTTGCCTCTGGTTGAAT
5'5OT3_Fw	TCTCATGCCTCCAGCACTATTTT
5'5OT3_Rv	TGATTGCCAGATTCAGGAGGAAA

9.1.4 Oligonucleotides used for the 3C assay

DpnII_1	AGGGGTTGCTGGAAAAGAAG
DpnII_4	GGCAGGATGGGACTGAAGTA
DpnII_5	ATTTGTCTGGGGGCTCATAAC
DpnII_8	CACACAACTATCACTACTATCACCAC

DpnII_10	TGTTGAATCCCACCTTTACTCC
DpnII_1B	CCATGCCCTGCTAAATGTGTA
DpnII_CNS2	TGTAGAAGGAAGGGCTCAGTGT
DpnII_dwCNS2	TGGAAAATGAGACACAACGAAG
DpnII_i1	AAGGACGGAGTGGAGGTTG
DpnII_i3	TTGGACCTGTGCTGTGACTAA
DpnII_C0	TCTTAATGTTGGAAAGTGCAAAGT
DpnII_C3	TCACAACACTGTCATACCATCTG
DpnII_D3	TTTTCTTGACTTTGTGTCCCTATG
DpnII_8i1	TTCTCTTTTCTTTCGCACCA
DpnII_8i3	AAAGACACCATCCCTCCAAC
DpnII_Grm5	CTATGGACATCAAAAATGCTAAGGT
DpnII_8CNS2	CCCAGACCCTTCCAAGTCAGTAT
DpnII_8Grm5	TACAGCAACACATTAGAACCAGA
DpnII_CNS2-3'	TCTCAAATCCCTCCTATCCAA
DpnII_CNS2-Tyrp	AGGCTGAGAGTATTTGATGTAAGAA
DpnII_XBP1	GTCTGTCTTTGTTGCTGAAGTATG
DpnII_XBP2	AAGTCCAGTGTGCTGAGGTATT

9.1.5 Oligonucleotides used for colony-PCR and sequencing of TALEN plasmids

pCR8_F1	TTGATGCCTGGCAGTTCCCT
pCR8_R1	CGAACCGAACAGGCTTATGT
TAL_F1	TTGGCGTCGGCAAACAGTGG
TAL_R2	GGCGACGAGGTGGTCGTTGG
seqTALEN5_1	TTGGCGTCGGCAAACAGTGG

9.1.6 Oligonucleotides used for CRISPR sgRNA cloning

LKO_1_5'	GACTATCATATGCTTACCGT
Tyr5'0gRNAFw	ACACCGTGTGACAGTGCAAGATAACG
Tyr5'0gRNARv	AAAACGTTATCTTGCACTGTCACACG
Tyr5'5gRNAFw	ACACCGTGATAAAACTAGGCAATTTG
Tyr5'5gRNARv	AAAACAAATTGCCTAGTTTTATCACG
Tyr3'3gRNAFw	ACACCAGCTCAGTAGAGTACTAGGTG
Tyr3'3gRNARv	AAAACACCTAGTACTCTACTGAGCTG
Tyr3'5gRNAFw	ACACCACAAACGCTAATTGGTAAAAG
Tyr3'5gRNARv	AAAACTTTTACCAATTAGCGTTTGTG

9.1.7 Oligonucleotides used for Cas9 and sgRNA in vitro transcription

Tyr50-FwJae	TTAATACGACTCACTATAGGGTGTGACAGTGCAAGATAAC
Tyr55-FwJae	TTAATACGACTCACTATAGGGTGATAAAACTAGGCAATTT
Tyr33-FwJae	TTAATACGACTCACTATAGGAGCTCAGTAGAGTACTAGGT
Tyr35-FwJae	TTAATACGACTCACTATAGGACAAACGCTAATTGGTAAAA
gRNA-RvJae	AAAAGCACCGACTCGGTGCC
T7_Cas9Fw	TAATACGACTCACTATAGGGAGAATGGACAAGAAGTACTCCATTG
Cas9Rv(T7)	CGGTAGGGATCGAACCCTTTCA

9.2: CRISPR predicted off-target sites

Comprehensive list of top-scoring off-target sites relative to the sgRNA used in this study, predicted using CRISPR Design online tool (crispr.mit.edu). Note that only a fraction of the predicted off-targets contains the PAM motif (NGG). The

genomic coordinates refer to the mouse genome assembly NCBI37/mm9 (July 2007).

9.2.1 The top 20 predicted off-targets of CRISPR 3'3

sequence	score	mismatch position	chromosomal coordinates
TGCTCAGGAGTGTACTAGGTTAG	1.6	3MMs [1:8:11]	chr12:+104596952
AGCTGGGCAGAGTACTAGGTTGG	1.4	3MMs [5:6:8]	chr11:+74783860
CTCTGAGTAAAGTACTAGGTTGG	1.3	4MMs [1:2:5:10]	chr11:-17927436
AGCTCAGGAGAGTCCTAGGTAGG	1	2MMs [8:14]	chr5:+135521999
AGCTCTGGTGAGTACTAGGTGGG	0.9	3MMs [6:8:9]	chr4:-122729022
AGCTTAAAATAGTACTAGGTCAG	0.8	4MMs [5:7:8:10]	chrX:-17285464
TGCTTAGTAGAGTAATAGGTGAG	0.8	3MMs [1:5:15]	chr5:-14857691
ATCTCAGTAGTTTACTAGGTCAG	0.8	3MMs [2:11:12]	chr17:+27459237
TGCCCAGTAATGTACTAGGTAAG	0.7	4MMs [1:4:10:11]	chr7:-150041150
AGATCATTAGAGAACTAGGTCAG	0.7	3MMs [3:7:13]	chr9:-63341008
ACATAAGTAGAGTACTAGGGAAG	0.7	4MMs [2:3:5:20]	chr5:+48853286
TACTCAGTACAGTACTAGGAAGG	0.7	4MMs [1:2:10:20]	chr17:+92410256
AGCAGAGGAGAGTACTGGGTAGG	0.6	4MMs [4:5:8:17]	chr6:-66223870
GACTCAGAAGAGAACTAGGTGGG	0.6	4MMs [1:2:8:13]	chr7:+51356271
GACTCAGAAGAGAACTAGGTGGG	0.6	4MMs [1:2:8:13]	chr7:+51306259
AGTGCACTAGTGTACTAGGTAGG	0.5	4MMs [3:4:7:11]	chr15:+97166248
ACCATAGTAGAGTACTAGCTGAG	0.5	4MMs [2:4:5:19]	chr13:-36182778
AGCAAAGTAGAGTACAAGGTGAG	0.5	3MMs [4:5:16]	chr16:-40186198
AGCACAATAAAATACTAGGTGGG	0.4	4MMs [4:7:10:12]	chr17:+61792828
ACGTCAGGAGAGTAGTAGGTCAG	0.4	4MMs [2:3:8:15]	chr18:+12340803

9.2.2. The top 20 predicted off-targets of CRISPR 3'5

sequence	score	mismatch position	chromosomal coordinates
ACAATTGGTAATTGGTAAAATGG	1.4	3MMs [5:6:8]	chr12:-36137724
ATACTCGATAATTGGTAAAACAG	1.4	4MMs [2:4:5:8]	chr3:-41928349
AGATTCGCTGATTGGTAAAAAGG	1.3	4MMs [2:4:5:10]	chr5:-42142945
ACTAACGATAATAGGTAAAAGAG	1.1	3MMs [3:8:13]	chr12:-95704167
ACAAATAGTAATTGGTAAAAAGG	1	3MMs [6:7:8]	chr15:-15417522
GCAGACAATAATTGGTAAAATAG	0.9	4MMs [1:4:7:8]	chr15:+24340490
CAAAACTATAATTGGTAAAAAAG	0.9	4MMs [1:2:7:8]	chr9:-95810584
AGTTACCCTAATTGGTAAAAAAG	0.9	4MMs [2:3:4:7]	chr2:+135082013
ATCAACTCTGATTGGTAAAAGAG	0.9	4MMs [2:3:7:10]	chr6:+140276804
ACATTCACTGATTGGTAAAATGG	0.9	4MMs [4:5:7:10]	chr12:-90377648
CCAAGAGGTAATTGGTAAAATAG	0.8	4MMs [1:5:6:8]	chr8:-30786668
TCATAAGCTCATTGGTAAAAGAG	0.8	4MMs [1:4:6:10]	chrX:+74303266
AAAAAAGTTTATTGGTAAAACGG	0.8	4MMs [2:6:8:10]	chr8:-9748974
ACAAACACTTATTGGTCAAACAG	0.7	3MMs [7:10:17]	chr2:+124209059
AGGGACGCTAATTGGTTAAATGG	0.6	4MMs [2:3:4:17]	chr19:-22527730
AAAAACTTTATTTGGTAAAACAG	0.5	4MMs [2:7:8:11]	chr6:+10580070
AAAAACTCAGATTGGTAAAAAAG	0.5	4MMs [2:7:9:10]	chr6:+141388187

TCAGAAGCAAATTGGTAAAAAAG	0.5	4MMs [1:4:6:9]	chr11:-8525080
GTAAATGCAAATTGGTAAAATGG	0.5	4MMs [1:2:6:9]	chr6:+105330140
AAAAACAGTAACTGGTAAAAAAG	0.5	4MMs [2:7:8:12]	chr17:+61574199

9.2.3 The top 20 predicted off-targets of CRISPR 5'0

sequence	score	mismatch position	chromosomal coordinates
CTGTCACAGAACAAGATAACCGG *	0.7	4MMs [1:5:10:11]	chr6:-99084014
GAGGGACAGACCAAGATAACCAG	0.7	4MMs [2:4:10:11]	chr13:-9319840
TAATGACAGTGCAAGATAAGTAG	0.7	4MMs [1:2:3:20]	chr1:-12346982
GGGAGACTGTGCAAGATAAGGGG *	0.7	4MMs [2:4:8:20]	chrX:+69682563
TTTTGACAGAGCAAGATAAATGG *	0.6	4MMs [1:3:10:20]	chr3:-60502350
GTGAGAGGCTGCAAGATAACAGG	0.6	4MMs [4:7:8:9]	chr15:+75075785
GTATGTCTTTGCAAGATAACTAG	0.5	4MMs [3:6:8:9]	chr10:-46557450
GTCTTAAAGTGGAAGATAACCAG	0.5	4MMs [3:5:7:12]	chr1:+165501357
GTCAGAGAGTGTAAGATAACAAG	0.5	4MMs [3:4:7:12]	chr1:-158207835
TTGTGATAGAGAAAGATAACAAG	0.5	4MMs [1:7:10:12]	chr17:+46477157
CTGTGAAAGAGAAAGATAACAAG	0.5	4MMs [1:7:10:12]	chr13:+23225282
GTGTGACAGAGAAAGAAAACCAG	0.5	3MMs [10:12:17]	chr14:-75695598
TTGTTACAATGGAAGATAACTAG	0.4	4MMs [1:5:9:12]	chr11:-37560558
ATGTGGCTGTGTAAGATAACTAG	0.4	4MMs [1:6:8:12]	chr4:-33292270
GTGGGATAGAGGAAGATAACCAG	0.4	4MMs [4:7:10:12]	chr4:-46935274
GGTTGACACTGGAAGATAACAAG	0.4	4MMs [2:3:9:12]	chr4:-95418322
GTTTGACAAAGTAAGATAACAAG	0.4	4MMs [3:9:10:12]	chr12:-7026399
GAGTATCAGTGCAAGACAACTAG	0.4	4MMs [2:5:6:17]	chr7:-149497895
GTGTGACCTGGCAAGATAAGCAG	0.4	4MMs [8:9:10:20]	chr13:+100741631
TTGTTACAGTCCAAGAAAACTGG	0.3	4MMs [1:5:11:17]	chr6:+59164176

* off-target sites analysed by T7 Endonuclease I assay

9.2.4 The top 20 predicted off-targets of CRISPR 5'5

sequence	score	mismatch position	chromosomal coordinates
GTCTTAAAACTAGGCAATTATGG	1.4	3MMs [3:4:20]	chr8:-34691408
GCACCAAAACTAGGCAATTTTAG	1.3	4MMs [2:3:4:5]	chr15:+9069056
GTATAAAAATTAGGCAATTTTAG	1.3	4MMs [3:4:5:10]	chr1:+143945358
ATGATCAAACAAGGCAATTTAAG	0.9	3MMs [1:6:11]	chr12:-75659983
GAAGTAAATCTAGGCAATTTGAG	0.8	4MMs [2:3:4:9]	chr12:+38489349
GGTAAGAAACTAGGCAATTTTAG	0.8	4MMs [2:3:5:6]	chr9:-84939931
GAGAATAAATTAGGCAATTTGAG	0.8	4MMs [2:5:6:10]	chr3:+138928190
GATATAACACGAGGCAATTTCAG	0.8	4MMs [2:3:8:11]	chr12:+70579083
CAGATAAAAGCAGGCAATTTTGG *	0.7	4MMs [1:2:10:11]	chr1:-9300003
GTGATGAAAGTAGGCAATTATAG	0.7	3MMs [6:10:20]	chr7:-17749834
GGAATAAAAAAAGGCAATTTAAG	0.7	4MMs [2:3:10:11]	chr9:-28015950
GAAATAAAAAAAGGCAATTTAAG	0.7	4MMs [2:3:10:11]	chr10:-129601638
GAGCAAAAACTTGGCAATTTGGG *	0.7	4MMs [2:4:5:12]	chr19:+31071409
GTATTAAAATGAGGCAATTTGGG *	0.7	4MMs [3:4:10:11]	chr14:+112763119
TTTATAAAAATAGGCAATTACAG	0.6	4MMs [1:3:10:20]	chr15:+47682160

ATGAAAAAAGTAGGCATTTTTGG	0.6	4MMs [1:5:10:17]	chr3:-27282906
GAGAAAAAAATAGGCATTTTTAG	0.6	4MMs [2:5:10:17]	chr14:+49087050
GTTTTCTAACTAGGCAATTTAGG	0.5	4MMs [3:4:6:7]	chr16:-45395464
TTGAGAAAATTAAGCAATTTAAG	0.5	4MMs [1:5:10:13]	chr9:-3459768
ATGTTAAAAATAAGCAATTTGGG	0.5	4MMs [1:4:10:13]	chr3:-47997909

* off-target sites analysed by T7 Endonuclease I assay